

IJCSIS Vol. 8 No. 6, September 2010
ISSN 1947-5500

International Journal of Computer Science & Information Security

© IJCSIS PUBLICATION 2010

Editorial Message from Managing Editor

IJCSIS is an open access publishing venue for research in general computer science and information security.

Target Audience: IT academics, university IT faculties; industry IT departments; government departments; the mobile industry and computing industry.

Coverage includes: security infrastructures, network security: Internet security, content protection, cryptography, steganography and formal methods in information security; computer science, computer applications, multimedia systems, software, information systems, intelligent systems, web services, data mining, wireless communication, networking and technologies, innovation technology and management.

The average paper acceptance rate for IJCSIS issues is kept at 25-30% with an aim to provide selective research work of quality in the areas of computer science and engineering. Thanks for your contributions in September 2010 issue and we are grateful to the experienced team of reviewers for providing valuable comments.

Available at <http://sites.google.com/site/ijcsis/>

IJCSIS Vol. 8, No. 6, September 2010 Edition

ISSN 1947-5500 © IJCSIS, USA.

Abstracts Indexed by (among others):



IJCSIS EDITORIAL BOARD

Dr. Gregorio Martinez Perez

Associate Professor - Professor Titular de Universidad, University of Murcia (UMU), Spain

Dr. M. Emre Celebi,

Assistant Professor, Department of Computer Science, Louisiana State University in Shreveport, USA

Dr. Yong Li

School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China

Prof. Hamid Reza Naji

Department of Computer Engineering, Shahid Beheshti University, Tehran, Iran

Dr. Sanjay Jasola

Professor and Dean, School of Information and Communication Technology, Gautam Buddha University

Dr Riktesh Srivastava

Assistant Professor, Information Systems, Skyline University College, University City of Sharjah, Sharjah, PO 1797, UAE

Dr. Siddhivinayak Kulkarni

University of Ballarat, Ballarat, Victoria, Australia

Professor (Dr) Mokhtar Beldjehem

Sainte-Anne University, Halifax, NS, Canada

Dr. Alex Pappachen James, (Research Fellow)

Queensland Micro-nanotechnology center, Griffith University, Australia

Dr. T.C. Manjunath,

ATRIA Institute of Tech, India.

TABLE OF CONTENTS

1. Paper 23081018: Improvement Dynamic Source Routing Protocol by Localization for Ad hoc Networks (pp. 1-6)

Mehdi Khazaei

Kermanshah University of Technology, Information Technology Engineering Group, Kermanshah, Iran

2. Paper 31081053: Steganalysis of Reversible Vertical Horizontal Data Hiding Technique (pp. 7-12)

Thom Ho Thi Huong, Faculty of Information Technology, Haiphong Private University, Haiphong, Vietnam

Canh Ho Van, Dept. of Professional Technique, Ministry of Public Security, Hanoi, Vietnam

Tien Trinh Nhat, College of Technology, Vietnam National University, Hanoi, Vietnam

3. Paper 21081012: Off-line Handwritten Signature Recognition Using Wavelet Neural Network (pp. 13-21)

Mayada Tarek, Computer Science Department, Faculty of Computers and Information Sciences, Mansoura, Egypt

Taher Hamza, Computer Science Department, Faculty of Computers and Information Sciences, Mansoura, Egypt

Elsayed Radwan, Computer Science Department, Faculty of Computers and Information Sciences, Mansoura, Egypt

4. Paper 21081013: A Black-Box Test Case Generation Method (pp. 22-31)

Nicha Kosindrdecha, Autonomous System Research Laboratory, Faculty of Science and Technology, Assumption University, Bangkok, Thailand

Jirapun Daengdej, Autonomous System Research Laboratory, Faculty of Science and Technology, Assumption University Bangkok, Thailand

5. Paper 21081014: White-Box Test Reduction Using Case-Based Maintenance (pp. 32-40)

Siripong Roongruangsuwan, Autonomous System Research Laboratory, Faculty of Science and Technology, Assumption University, Bangkok, Thailand

Jirapun Daengdej, Autonomous System Research Laboratory, Faculty of Science and Technology, Assumption University, Bangkok, Thailand

6. Paper 23081021: Nat Traversal for Video Streaming Applications (pp. 41-46)

*Omar A. Ibraheem^{#1}, Omer Amer Abouabdalla^{*2}, Sureswaran Ramadass^{#3}*

[#] National Advanced IPv6 center of Excellence (NAV6), Universiti Sains Malaysia (USM), Pulau penang, Malaysia

7. Paper 25081026: The Integration of GPS Navigator Device with Vehicles Tracking System for Rental Cars Firms (pp. 47-51)

Omarah O. Alharaki, KICT, International Islamic University, Kuala Lumpur, Malaysia

Fahad S. Alaieri, KICT, International Islamic University, Kuala Lumpur, Malaysia

Akram M. Zeki, KICT, International Islamic University, Kuala Lumpur, Malaysia

8. Paper 29081036: Process Framework in Global eXtreme Programming (pp. 52-59)

Ridi Ferdiana, Lukito Edi Nugroho, Paulus Insap Santoso

Department of Electrical Engineering and Information Technology, Gadjah Mada University (UGM)

Yogyakarta, Indonesia

Ahmad Ashari, Department of Computer Science and Electronics, Gadjah Mada University (UGM), Yogyakarta, Indonesia

9. Paper 29081038: A Hybrid PSO-SVM Approach for Haplotype Tagging SNP Selection Problem (pp. 60-65)

Min-Hui Lin, Department of Computer Science and Information Engineering, Dahan Institute of Technology, Sincheng, Hualien County 971, Taiwan, Republic of China
Chun-Liang Leu, Department of Information Technology, Ching Kuo Institute of Management and Health, Keelung 336, Taiwan, Republic of China

10. Paper 07110912: PAPR Reduction Technique for LTE SC-FDMA Systems Using Root-Raised Cosine Filter (pp. 66-71)

Md. Masud Rana, Jinsang Kim and Won-Kyung Cho
Department of Electronics and Radio Engineering, Kyung Hee University
1 Seocheon, Kihung, Yongin, Gyeonggi, 449-701, Republic of Korea

11. Paper 22081017: Survey of Routing Protocols and Channel Assignment protocols in Wireless Mesh Networks (pp. 72-77)

Vivek M Rathod, Suhas J Manangi, Satish E, Saumya Hegde
National Institute of Technology Karnataka – Surathkal

12. Paper 29071044: An Approach For Designing Distributed Real Time Database (pp. 78-87)

Dr. Dhuha Basheer Abdullah, Computer Sciences Dept./Computers Sciences and Mathematics College /Mosul University, Mosul- Iraq
Ammar Thaher Yaseen, Computer Sciences Dept./Computers Sciences and Mathematics College /Mosul University, Mosul- Iraq

13. Paper 31081051: Person Identification System using Static-dynamic signatures fusion (pp. 88-92)

Dr. S.A Daramola¹ and Prof.T.S Ibiyemi²
¹Department of Electrical and Information Engineering, Covenant University Ota Ogun State, Nigeria.
²Department of Electrical Engineering, University of Ilorin, Ilorin, Kwara-State, Nigeria

14. Paper 31081061: Short term flood forecasting using RBF static neural network modeling a comparative study (pp. 93-98)

Rahul P. Deshmukh, Indian Institute of Technology, Bombay, Powai, Mumbai, India
A. A. Ghatol, Former Vice-Chancellor, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, India

15. Paper 31031083: Analysis of impact of Symmetric Encryption Algorithms in Data Security Model of Grid Networks (pp. 99-106)

N. Thenmozhi, Department of Computer Science, N.K.R. Govt. Arts College for Women, Namakkal-637 001, India.
M. Madheswaran, Department of Electronics and Communication Engg., Muthayammal Engineering College, Rasipuram-637 408, India.

16. Paper 09071005: Low Power and Area Consumption Custom Networks-On-Chip Architectures Using RST Algorithms (pp. 107-115)

1 P.Ezhumali 2 Dr.C.Arun
1 Professor, Dept of Computer Science Engineering
2 Asst. Professor, Dept of Electronics and Communication
Ralakshmi Engineering College, Thandalam-602 105, Chennai, India

17. Paper 16081005: Prediction of Epileptic form Activity in Brain Electroencephalogram Waves using Support vector machine (pp. 116-121)

Pavithra Devi S T, M.Phil Research Scholar, PSGR Krishnammal College for Women, Coimbatore Tamilnadu, India
Vijaya M S, Assistant Professor and Head, GRG School of Applied Computer Technology, PSGR Krishnammal College for Women, Coimbatore Tamilnadu, India

18. Paper 16081006: Deployment of Intelligent Agents in Cognitive networks (pp. 122-127)

Huda Fatima, Dept. of CS, Jazan University, Jazan, K.S.A

Dr. Sateesh Kumar Pradhan, Dept. of Comp.Engineering, King Khalid University, Abha, K.S.A

Mohiuddin Ali Khan, Dept. of Comp. Networks, Jazan University, Jazan, K.S.A

Dr. G.N.Dash, Dept. of Comp. Science, Sambalpur University, Orissa, India

19. Paper 21081011: A Performance Study on AES Algorithms (pp. 128-133)

B.D.C.N.Prasad¹, P.E.S.N.krishna Prasad², P Sita Rama Murty³ and K Madhavi⁴

1. Dept. of Computer Applications, P V P Siddardha Institute of Technology, Vijayawada,

2. Dept. of CSIT, Sri Prakash College of Engineering, Tuni,

3. Dept. of CSIT, Sri Prakash College of Engineering, Tuni,

4. Dept. of CSE, Dadi Institute of Technology, Anapalli,

20. Paper 24091018: Hybrid Fingerprint Image compression and Decompression Technique (pp. 134-138)

Dr.R.Seshadri, ,B.Tech.,M.E,Ph.D, Director, S.V.U.Computer Center S.V.University, Tirupati

Yaswanth Kumar.Avulapati , M.C.A, M.Tech, (PhD), Research Scholar, Dept of Computer Science, S. V. University, Tirupati

Dr.M.Usha Rani M.C.A, PhD, Associate Professor, Dept. of Computer Science,, SPMVV, Tirupati

21. Paper 25081025: Punctured Self-Concatenated Trellis Codes with Iterative Decoding (pp. 139-144)

Labib Francis Gergis, Misr Academy, Mansoura City, Egypt

22. Paper 26081028: Application of Fuzzy Composition Relation For DNA Sequence Classification (pp. 145-148)

Amrita Priyam, Dept. of Computer Science and Engineering, Birla Institute of Technology, Ranchi, India.

B. M. Karan⁺, G. Sahoo⁺⁺

⁺Dept. of Electrical and Electronics Engineering, ⁺⁺Dept. of Information Technology

Birla Institute of Technology, Ranchi, India

23. Paper 26081029: Data Security in Mobile Ad Hoc Networks using Genetic Based Biometrics (pp. 149-153)

B. Shanthini, Research Scholar , CSE Department , Anna University , Chennai, India

S. Swamynathan, Assistant Professor, CSE Department , Anna University , Chennai, India

24. Paper 28081031: Effective Multi-Stage Clustering for Inter- and Intra-Cluster Homogeneity (pp. 154-160)

Sunita M. Karad[†], Assistant Professor of Computer Engineering, MIT, Pune, India

V.M.Wadhai^{††}, Professor and Dean of Research, MITSOT, MAE, Pune, India

M.U.Kharat^{†††}, Principle of Pankaj Laddhad IT, Yelgaon, Buldhana, India

Prasad S.Halgaonkar^{††††}, Faculty of Computer Engineering, MITCOE, Pune, India

Dipti D. Patil^{†††††}, Assistant Professor of Computer Engineering, MITCOE, Pune, India

25. Paper 28081033: A Pilot Based RLS Channel Estimation for LTE SC-FDMA in High Doppler Spread (pp. 161-166)

M. M. Rana

Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology, Khunla, Bangladesh

26. Paper 28081034: Priority Based Congestion Control for Multimedia Traffic In 3G Networks (pp. 167-173)

Prof V.S Rathore 1, Neetu Sharma 2, Amit Sharma 3, Durgesh Kumar Mishra 4

123 Department of Computer Engineering, Rajasthan, India

12 Rajasthan College of Engineering for women, Rajasthan, India

3 Shri Balagi College of Engineering & Technology, Rajasthan, India

4 Acropolis Institute of Technology and Research, Indore, MP, India

27. Paper 31081050: Adaptive Sub-block ARQ techniques for wireless networks (pp. 174-178)

A. N. Kemkar, Member, ISTE and Dr. T. R. Sontakke, Member,ISTE

28. Paper 30071074: Trigon-based Authentication Service Creation with Globus Middleware (pp. 179-185)

Ruckmani V , Ramakrishna Engineering College, Coimbatore,India

Anitha Kumari K , PSG College of Technology, Coimbatore,India

Sudha Sadasivam G , PSG College of Technology, Coimbatore,India

Dhaarini M P , PSG College of Technology, Coimbatore,India

29. Paper 30081040: Performance Evaluation of Speaker Identification for Partial Coefficients of Transformed Full, Block and Row Mean of Speech Spectrogram using DCT, WALSH and HAAR (pp. 186-198)

Dr. H. B. Kekre, Senior Professor, MPSTME, SVKM's NMIMS University, Mumbai, 400-056, India

Dr. Tanuja K. Sarode, Assistant Professor, Thadomal Shahani Engg. College, Bandra (W), Mumbai, 400-050, India

Shachi J. Natu, Lecturer, Thadomal Shahani Engg. College, Bandra (W), Mumbai, 400-050, India

Prachi J. Natu, Assistant Professor, GVAIET, Shelu, Karjat 410201, India

30. Paper 30081041: A Research Proposal for Mitigating DoS Attacks in IP-based Networks (pp. 199-201)

Sakharam Lokhande, Assistant Professor, School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606.

*Parag Bhalchandra **, Assistant Professor, School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606.*

Nilesh Deshmukh, Assistant Professor , School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606.

Dr. Santosh Khamitkar, Assistant Professor , School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606.

Santosh Phulari, Assistant Professor, School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606.

Ravindra Rathod, Assistant Professor, School of Computational Science, Swami Ramanand Teerth Marathwada University, Nanded, MS, India, 431606

31. Paper 30081042: An Efficient and Minimum Cost Topology Construction for Rural Wireless Mesh Networks (pp. 202-209)

Prof. V. Anuratha, H.O.D – PG. Comp. Science, Sree Saraswathi Thyagaraja college, Pollachi, Tamil Nadu, India

Dr. P. Sivaprakasam , Associate Professor, Sri Vasavi college of Arts & Science, Erode, Tamil Nadu, India

32. Paper 31071081: Reinforcement Learning by Comparing Immediate Reward (pp. 210-214)

Punit Pandey, Department Of Computer Science and Engineering, Jaypee University Of Engineering And Technology

Dr. Shishir Kumar, Department Of Computer Science and Engineering, Jaypee University Of Engineering And Technology

Deepshikha Pandey, Department Of Computer Science and Engineering, Jaypee University Of Engineering And Technology

33. Paper 31081043: Information realization with statistical predictive inferences and coding form (pp. 215-220)

D. Mukherjee, Sir Padampat Singhania University, Udaipur-313601,Rajasthan,India

P.Chakrabarti, A.Khanna , V.Gupta*

Sir Padampat Singhania University, Udaipur-313601,Rajasthan,India

34. Paper 31081045: Scaling Apriori for Association Rule Mining using Mutual Information Based Entropy (pp. 221-227)

S. Prakash, Research Scholar, Sasurie College of Engineering, Vijayamangalam, Erode(DT), Tamilnadu, India.

Dr. R. M. S. Parvathi M.E.(CSE), Ph.D., Principal, Sengunthar College of Engg. for Women, Tiruchengode. Tamilnadu, India.

35. Paper 31081046: Clustering of High-Volume Data Streams In Network Traffic (pp. 229-233)

M. Vijayakumar, Research Scholar, Sasurie College of Engineering, Vijayamangalam, Erode(Dt) , Tamilnadu, India.

Dr. R.M.S. Parvathi M.E.(CSE), Ph.D., Principal, Sengunthar College of Engg. for Women, Tiruchengode. Tamilnadu, India.

36. Paper 31081049: A.R.Q. techniques using Sub-block retransmission for wireless networks (pp. 234-237)

A. N. Kemkar, Member, ISTE and Dr. T. R. Sontakke, Member, ISTE

37. Paper 22081015: Performance Analysis of Delay in Optical Packet Switching Using Various Traffic Patterns (pp. 238-244)

A. Kavitha , IT dept, Chettinad College of Engineering & Technology, Karur, Tamilnadu, India

V. Rajamni, Indra Ganesan College of Engineering, Trichy, Tamilnadu, India

P. Anandhakumar, IT Dept, Madras Institute of Technology, Chennai, Tamilnadu, India

38. Paper 29081039: A Feedback Design for Rotation Invariant Feature Extraction in Implementation with Iris Credentials (pp. 245-254)

M. Sankari, Department of Computer Applications, Nehru Institute of Engineering and Technology, Coimbatore, India.

R. Bremananth, School of EEE, Information Engg. (Div.), Nanyang Technological University, Singapore - 639798.

39. Paper 31081047: Empirical Mode Decomposition Analysis of Heart Rate Variability (pp. 255-258)

C. Santhi, M.E., Assistant Professor, Electronics and Communication Engineering, Government College of Technology, Coimbatore-641 013

N. Kumaravel, Ph.D, Professor, Head of the Department, Electronics and Communication Engineering, Anna University, Chennai-600 025.

Improvement Dynamic Source Routing Protocol by Localization for Ad hoc Networks

Mehdi Khazaei
Kermanshah University of Technology
Information Technology Engineering Group
Kermanshah, Iran .

Abstract-Ad hoc networks are temporary networks with a dynamic topology which don't have any established infrastructure or centralized administration. Consequently, in recent years many researchers have focused on these networks. These networks need efficient routing protocols in terms of Quality of Services (QoS) metrics. Ad hoc networks suffer from frequent and rapid topology changes that cause many challenges in their routing. Most of the routing protocols like this proposed protocol try to find a route between source and destination nodes and when any route is expired, a new route would be formed. Rapid route reconstruction may cause the network inefficiency. Therefore, we have to decrease this processes. The proposed protocol as DSR routing protocol build one routes between source and destination but create backup routes during the route reply process, route maintenance process and use local recovery process in order to improve the data transfer and attended to QoS. The protocol performance is demonstrated by using the simulation results obtain from the global mobile simulation software (Glomosim). The experimental results show that this protocol can decrease the packet loss ratio and increase data transfer rather than DSR that, it is useful for the applications that need a high level of reliability.

Keywords; *Protocol, Routing, Local Recovery, Mobile Ad-hoc Networks*

I. Introduction

Routing in ad hoc networks is a very challenging issue due to nodes mobility, dynamic topology, frequent link breakage, limitation of nodes (memory, battery, bandwidth, and processing power), and limited transmission range of the node and lack of central point like base stations or servers. On the other hand, there are a lot of performance metrics and quality services which should be satisfied in an ad hoc network like end -to-end data throughput, average end-to-end data delay, packet loss ratio, Normalized Routing Load, Packet Delivery Ratio, and route optimality. Each protocol can satisfy some of these metrics and has some drawbacks in terms of other metrics. Furthermore, due to the nature of ad hoc networks (distributed and cooperated routing), even for a fixed metric, each protocol can show a different performance with different networks features like number of mobile nodes, mobility of nodes, pause time and.... So by comparing between different ad hoc routing protocols we can extract very important information about the performance of these protocols in the Different situations. In the other hands, the nodes mobility and the probability of links failure may cause the fault tolerance

Issues more important for routing problem in ad hoc network therefore, each routing protocol should be fault tolerant in probable route failures [1].

Routing protocols in conventional wired networks are usually based upon either distance vector or link state routing algorithms as a DSDV [2], CGSR [2] and FSR [2]. These algorithms require periodic routing advertisements to be broadcast by each router. These conventional routing algorithms are clearly not efficient for type of dynamic changes which may occur in an ad-hoc network [2, 3]. A new class of on-demand routing protocols e.g., DSR [4, 5], TORA [2], AODV [6, 7]) for mobile ad hoc networks has been developed with the goal of minimizing the routing overhead. These Protocols reactively discover and maintain only the needed routes, in contrast to proactive protocols (e.g., DSDV [2]) which maintain all routes regardless of their usage. The key characteristic of an on-demand protocol is the source-initiated route discovery procedure. Whenever a traffic source needs a route, it initiates a Route discovery process by sending a route request for the destination (typically via a network-wide flood) and Waits for a route reply. Each route discovery flood is Associated with significant latency and overhead. This is particularly true for large networks. Therefore, for on-demand routing to be effective, it is desirable to keep the route discovery frequency low [8].

Single route routing allows the establishment of one route between a source and single destination node. Because of node mobility, the route may be broken frequently; therefore, having replacement route in cache memory to transmit data will improve the fault tolerance and higher aggregate bandwidth in these networks. Beside of this, by repairing the broken routes locally, the number of route rediscovery processes can be decreased. This paper improves the fault tolerance and increase reliability by obtain replacement routes in RREP¹ and RRER² processes and local recovery process together. This optimization is done on the DSR protocol.

The rest of this paper is organized as follows. In section II the DSR protocol is explained. Section III deals with the related works and Section IV describe the proposed protocol mechanism in detail. Performance evaluation by simulation is presented in section V and concluding remarks are made in section VI.

¹Route reply

²Route error

II. Dynamic Source Routing Protocol (DSR)

DSR is an on-demand routing protocol for ad hoc networks. Like any source routing protocol, in DSR the source includes the full route in the packets' header. The intermediate nodes use this to forward packets towards the destination and maintain a route cache containing routes to other nodes. In following subsections DSR operation are briefly described [4].

A. Route discovery

If the source does not have a route to the destination in its route cache, it broadcasts a RREQ³ message specifying the destination node for which the route is requested. The RREQ message includes a route record which specifies the sequence of nodes traversed by the message. When an intermediate node receives a RREQ, it checks to see if it is already in the Route record. If it is, it drops the message. This is done to prevent routing loops. If the intermediate node had received the RREQ before, then it also drops the message. The intermediate node forwards the RREQ to the next hop according to the route specified in the header. When the destination receives the RREQ, it sends back a route reply message. If the destination has a route to the source in its route cache, then it can send a RREP message along this route. Otherwise, the RREP message can be sent along the reverse route back to the source. Intermediate nodes may also use their route cache to reply To RREQs. If an intermediate node has a route to the destination in its cache, then it can append the route to the route record in the RREQ, and send an RREP back to the source containing this route. This can help limit flooding of The RREQ. However, if the cached route is out-of-date it can result in the source receiving stale routes [4].

B. Route maintenance

When a node detects a broken link while trying to forward a packet to the next hop, it sends a RERR message back to the source containing the link in error. When an RERR message is received, all routes containing the link in error are deleted at that node [4].

III. Related Works

Ad hoc routing protocols such as ADOV, DSR, DSDV and OLSR have been investigated on the ad hoc networks in the past few years. The investigations of the performance of these protocols on the ad hoc networks have produced many useful results. However, we have seen very limited findings of how these Ad-hoc routing protocols perform on wireless ad hoc networks. Nonetheless, we can see many attempts at developing routing protocols for ad hoc networks under the different deployment of ad hoc networks [8, 9 and 10]. In following are brought some of these attempts.

SMR is an on demand routing protocol that uses maximally disjoint routes to transmit data packets. Unlike DSR, intermediate nodes do not allow to send RREP packets back to the source instead, only destination nodes reply to the RREQ packets and selects maximally-disjoint routes [11].

MP-DSR is a multi-route QOS aware extension to DSR. It focuses on a QOS metric, end-to-end reliability. End-to-end

Reliability is defined as the probability of sending data successfully from the source to the destination node within a time window. MP-DSR selects a set of routes that satisfy a specific end-to-end reliability Requirement [12].

MSR is attempts to minimize the end-to-end delay for sending a data from source to destination by using multi-route routing and intelligent traffic allocation [13].

CHAMP is multi-route protocol that uses round-robin traffic allocation to keep routes fresh. It also employs cooperative packet caching to improve fault tolerance and takes advantage of temporal locality in routing, where a dropped packet is a recently sent packet [9].

The local recovery techniques have been used in some routing protocols for route maintenance processes. This technique aims to reduce the frequency of RREQ floods triggered by nodes that are located in the broken routes [14].

SLR is one of these routing protocols. It modifies DSR, using a new route recovery mechanism called bypass routing. Bypass routing utilizes both route caches and local error recovery techniques during failures to reduce the control overhead [15].

LRR is also another routing protocol that uses local recovery techniques. In this protocol the information of next-to-next (NN) node is stored at each intermediate node along the route. After detecting a broken link by an upstream node, it sends out the non-propagating requests to find another node which is in contact with itself and the NN node on the route; therefore the routes can be repaired locally in the shortest possible time [16].

MRFT protocol improves fault tolerance in DSR and SMR protocols. To achieve the goal of decreasing the packet loss ratio and increasing fault-tolerance, MRFT uses both multi-route and local recovery techniques together [17].

IV. The Proposed Protocol

This paper proposes IM-DSR⁴ protocol to improve fault tolerance and QOS in DSR protocol. To achieve the goal of decreasing the packet loss ratio and increasing fault-tolerance, IM-DSR uses local recovery techniques and alternate route during route reply and route maintenance that reliability in the network would be increased. IM-DSR modifying the route discovery, route reply and route maintenance processes in DSR. The IM-DSR protocol is including route discovery, route reply, route maintenance and local recovery processes that discussed in the following subsection.

A. Route Discovery Process

IM-DSR is an on demand routing protocol that builds single route using request/reply cycles. When the source node needs to send data to the destination but no route information is known, it floods RREQ packets over the entire network. When an intermediate node receives a RREQ that is not a duplicate, it appends its ID to the packet and rebroadcasts it. In IM-DSR all of the duplicate RREQs that are received by intermediate nodes are dropped. In IM-DSR, intermediate

³Route Request

⁴Improvement-DSR

Nodes are allowed to send RREPs back to the source even when they have route information to the destination in their route caches.

B. Route Reply Process

When receiving the first RREQ, the destination sends a RREP back to the source. After that, the destination node consumes other RREQs. The Route-Number of the RREP is one.

After receiving RREP packet by the intermediate nodes, if it has not route with same length to destination node, they store the routes in their route caches. The Route-Number of this routes are zero and used in route maintenance process for improving the break routes also sending data if there is not main route.

Look at the Fig. 1, suppose that node (H) sends the RREP to the source node (A), a route is found and sent to node (A) by RREP is A-C-D-G-F-H. Now, suppose that the RREP is received by node (C) which is middle node. Node (C) saves the routes to destination (H) which is C-D-G-F-H, additionally node (C) save C-D, C-D-G and C-D-G-F routes in the route caches.

When the source node receive the RREP, it will store the route and use that for transmit data.

C. Route Maintenance and Local Recovery Processes

During a transmission session, a problem such as node mobility, or low battery power might be raised, which can lead to break an existing route and lose route connectivity. This may force a route rediscovery process by flooding RREQs over the network. To avoid this phenomenon, IM-DSR uses following mechanism that one of them is local recovery techniques. Using local recovery techniques is very useful despite they consume the limited power of each nodes.

Suppose that a node finds a broken link, while sending a packet. At first, it seeks the route cache and deletes all routes include the broken link, and then according to kind of the packet one of the following items is done:

- If transitional packet would be a RREQ, the node would not send RRER to the source node.
- If transitional packet would be a RREP, send RRER to the node which makes the RREP.

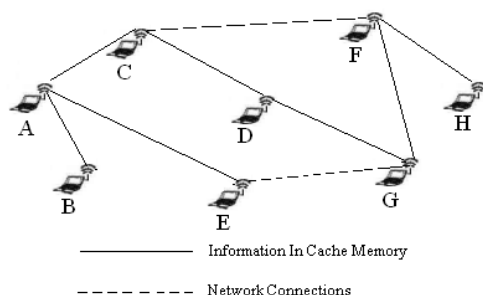


Figure 1. Routes Structure in an Ad-hoc Network

- If transitional packet is a RRER, it would examine how many times the packet would be saved. If it was the first time, the meaning is, the packet would be saved by examining a route cache and finding alternate route. The RRER is sent to destination through that route then the RRER is made and it will report the broken link to the source of RRER. If it were not first time or if alternate route were not in the route cache of node, the RRER would be deleted and only a RRER would be sent to the source node. Therefore RRER is saved only for one time by the IM-DSR protocol.
- If transitional packet would be data, it would examine how many times the packet would be saved. If this time were less than three, the data packet would be sent by examining their route cache and alternate route then it will send a RRER to the source node. If these times were more than three or if alternate route was not in the route cache of node the data packet would be deleted and only a RRER would be sent to the source node. Therefore data packet is saved for three times by the IM-DSR protocol.

If very data packet passed the same route towards destination node and they faced the broken link (while sending), the node which recognized an error, for every data packet send a RRER to the source node. In order to avoid this item every node before sending RRER to source node, examine this is a first RRER or not. If it was not send, a new RRER send to source node.

Every node which recognizes the broken link and makes the RRER, examined the route cache in order to find alternate route and put it in the RRER, which the node that received the RRER, replaces the route in the RRER with the previous invalid route in the route cache.

Fig. 2 shows this matter. The source node (A) sends data to the destination node (H) through A-C-D-H. When node (D) sends the data packet, it will find the failure in node (H). By examine the route cache, it chooses the alternate route D-G-F-H, hence, the data packet is sent to destination node (H) through this route, then RRER is made and it is sent to the node (A). This packet includes the alternate route D-G-F-H. The node (A) receives the RRER and deletes A-C-D-H from route cache and replaces A-C-D-G-F-H.

Every middle and source node which receives the RRER, examine those route in route cache which includes the broken link and should be deleted from cache and if packet included alternate route, exploited that and saved in route cache with number two. If in buffer, data packet waiting to send toward alternate route destination, it will send through that route. Such as the Fig. 2, while passing the RRER, node (C) adds C-D-G-F-H to the route cache.

If a node who detected a broken link cannot find any alternate route in its route cache, so it drops the data packet and sends a RERR without any repaired route to the source. After that, because of performing local recovery process by the node that detects the broken link, the source node does not trigger the rediscovery process immediately. After detecting the broken link, node sends a RERR to the source and starts the local recovery process simultaneously. To repair the route

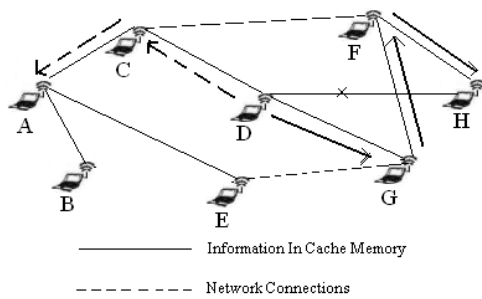


Figure 2. Ad-hoc Network Structure, when occur break link

Locally, node triggers a local query to its neighbors. The neighbors reply if they have any valid route to the destination. When node receives the RRP⁵, it repairs the primary route and then sends a RRP back to the source, like shown in the Fig. 3.

In Fig. 3 when node (D) finds the failure in node (G), First examined the route cache for replacement route, if not found then begun local recovery request which node (F) reply with F-G-H as repaired route. When node (D) receives this route update its memory cache and sends data from this route to destination and send RRP to the source node that here is node (A). When node (A) receives the RRP deletes A-C-D-G-H from route cache and replaces A-C-D-F-G-H. Every middle node which receives RRP, examine if included repaired route, then exploited that and saved in route cache with number two. In Fig. 3 while passing RRP, node (C) adds C-D-F-G-H to the route cache more node (C) adds C-D, C-D-F and C-D-F-G to route cache.

If the source node has not primary route to send data, it will use the repaired route. If no data packet containing repaired route was reported to the source node for a certain amount of time, then it sends a new RREQ to the destination node. After receiving the new RREQ by the destination node, it performs the route rediscovery process that was described in subsection A.

In the IM-DSR protocol, route with Route-Number one is main route. Route with Route-number zero are obtain by route reply process and route with Route-Number two are those which are obtain in route maintenance process and known as repaired route.

When the source node wants to send data to a destination, it tries to use the primary, which has the highest priority (the routes with route-Number one) at first, if not exist, it will use obtain route (the route with Route-Number zero) otherwise use repaired route (the route with Route-Number two) to send data.

V. Performance Evaluation

A. Simulation Environment

In order to demonstrate the effectiveness of IM-DSR protocol, we evaluate our proposed protocol and compare its performance to the DSR (uni-route). We have implemented IM-DSR protocol using the Global Mobile Simulation library

⁵ Route Repaired Packet

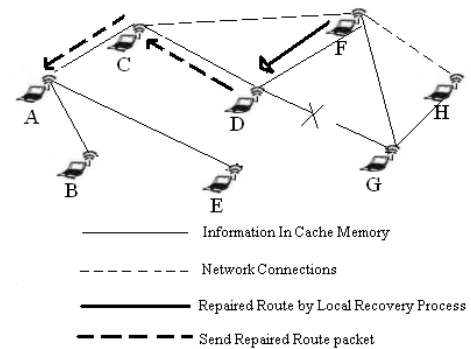


Figure 3. Ad-hoc Network Structure, when Occur Local Recovery

(GLOMOSIM) [18]. The Simulation environment consists of 50 numbers of nodes in a rectangular region of size 1500 meters by 1500 meters.

The nodes are randomly placed in the region and each of them has a radio propagation range of 250 meters. 200 constant bit rate (CBR) flows are deployed for data transmission. Simulation time is 300 seconds. The random waypoint model is chosen as the node mobility model. All data packets are 512 bytes. Band width is 2 Mbps and simulation done for 0, 1, 3, 5 and 10 second as stop time. Minimum and maximum speed for nodes are 0 m/s and 30 m/s. IEEE 802.11 selected for MAC layer protocol.

B. Simulation Results

Fig. 4 shows packet delivery ratio (PDR) in every two protocol. It is defined as ratio of the number of data packets delivered to the destinations generated by sources. In DSR protocol if sending node of data has not any routes to send data, it would start route request process by RREQ. In every request finds a route and if it was invalid, the route request process begins again. The IM-DSR protocol against DSR protocol obtains routes in route reply and Maintenance process and used them for sending data. This would lead to increase the packet delivery ratio in IM-DSR protocol.

Fig. 5 shows the number of RREQ (NRQ) in every two protocols. This number is sum of the RREQ in request process. Fig. 5 shows that this number in DSR protocol is greater than IM-DSR protocol because in every route discovery find only one route and if this route would be invalid begin route discovery again. But IM-DSR finds routes in route reply, maintenance and local recovery processes, hence reduces this number.

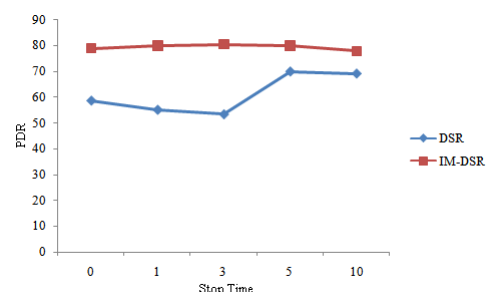


Figure 4. Packet Delivery Ratio (PDR)

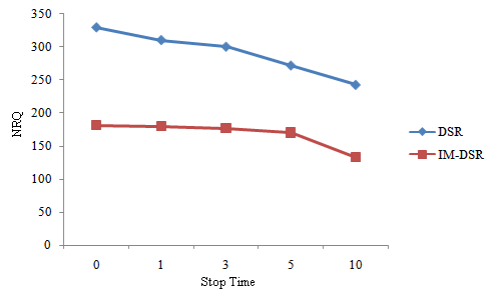


Figure 5. Number of RREQ (NRQ)

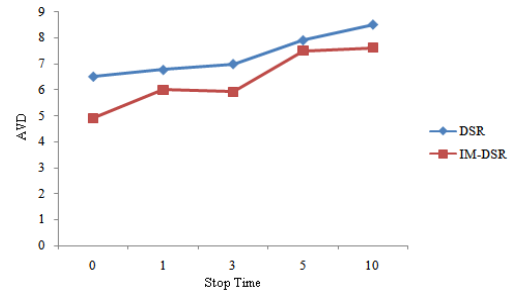


Figure 8. Average of Delay (AVD)

Fig. 6 shows sum of the hop-count (SHC) in the network. This number is summation of the all routes hop count that lower value leads to less delay in network. This number is less in IM-DSR protocol because of less route request process and found optimal routes.

Fig. 7 shows the number of broken links (NBL) in the network. Problem such as node mobility, low battery power or congestion might be raised, which can lead to break an existing route and lose route connectivity. Number of broken link in IM-DSR protocol is less than DSR protocol because of used alternate routes that making in route request, reply and maintenance processes and used local recovery process.

Fig. 8 shows an average of delay (AVD) in two protocols. Because of sending data from route with long hop-count and data packets waiting more time in buffer, delay in DSR protocol is more than IM-DSR protocol. The IM-DSR protocol which, between the source and destination, selected optimal routes and saved multi-route node decreases the delay in comparisons with DSR protocol.

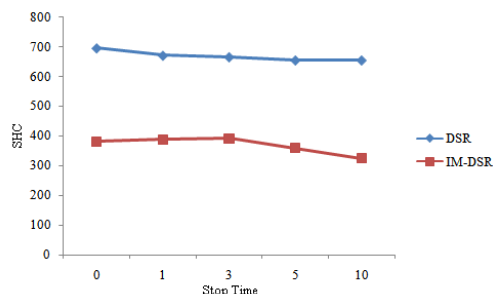


Figure 6. Sum of Hop-Count (SHC)

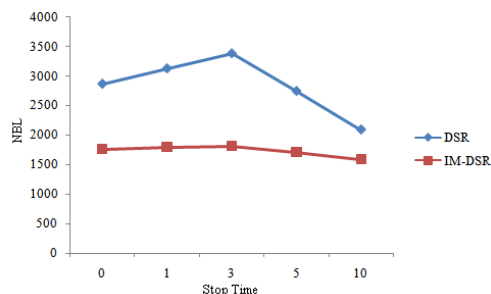


Figure 7. Number of Broken Link (NBL)

VI. Conclusion

In this paper, we proposed a new routing protocol called IM-DSR to provide higher QoS in ad hoc networks. This protocol is an extension of DSR to increase the reliability by modifying the route discovery and route maintenance processes in DSR also added the local recovery techniques to DSR. The simulation results show that IM-DSR is very effective in decreasing the packet loss and also increasing the fault tolerance ad hoc networks. In all of the cases, IM-DSR has the higher packet delivery ratio than the DSR protocol while improving the overhead of route maintenance, maintaining acceptable overhead. Therefore the proposed routing protocol is very useful for the applications that need a high level of reliability.

REFERENCES

- [1] V. Nazari Talooki, J. Rodriguez, "Quality of Service for Flat Routing Protocols in Mobile Ad hoc Networks", Mobimedia'09, London, UK, September, 2009.
- [2] E. M. Royer, "A Review of current Routing Protocols for Ad Hoc Mobile Wireless Networks", IEEE Personal Communications, vol. 6, no. 2, pp. 46-55, 1999.
- [3] M. IILyas, "The handbook of ad hoc wireless networks", CRC press, 2003.
- [4] D. Johnson, Et. al, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Network" Internet Draft, www.ietf.org/internetdrafts/ draft-ietf-manet-dsr-10.txt (July 2004).
- [5] D. B. Johnson, D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks", Mobile Computing, Edited by Imielinski, Korth, Chapter 5, pp. 153-181, Kluwer Academic Publishers, 1996.
- [6] C. Perkins, E. Royer, "Ad Hoc On-Demand Distance Vector (AODV) Routing" (Internet Draft), www.ietf.org/internet-drafts/draft-ietfmanet-aodv-13.txt (Feb 2003).
- [7] C. Perkins, E. Royer, "Ad-hoc On-demand Distance Vector Routing", 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, United States, pp. 90-100, Feb. 1999.
- [8] Q. Jiang, D. Manivannan, "Routing protocols for Sensor networks", IEEE, pages 93-98, 2004.
- [9] S. Mueller, R.P. Tsang and D. Ghosal, "Multipath Routing in Mobile Ad hoc Networks Issues and Challenges", Proceedings of CA USA,27,USA,2007.
- [10] K. Wu, J. Harms, "On-Demand Multipath Routing for Mobile Ad Hoc Networks", 4th European Personal Mobile Communications Conference (EPMCC 01), Vienna, Austria, Feb. 2001.
- [11] S. Ju-Lee, M. Gerla, "Split Multipath Routing with Maximally Disjoint Paths in Ad Hoc Networks", IEEE International Conference on Communications, vol. 10, pp. 3201-3205, Helsinki, Jun. 2001.
- [12] R. Leung, E. Poon, A. C. Chan, B.Li, "MP-DSR: A QoS-Aware Multi path Dynamic Source Routing Protocol for Wireless Ad-Hoc Networks",

- 26th Annual IEEE International Conference on Local Computer Networks (LCN2001), pp. 132-141, Tampa, FL, United States, Nov. 2001.
- [13] X. Li, L. Cuthbert, "On-Demand Node-Disjoint Multipath Routing in Wireless Ad Hoc Networks", 29th Annual IEEE International Conference on Local Computer Networks (LCN2004), Tampa, FL, United States, pp. 419-420, Nov. 2004.
- [14] E. Esmaili, P. Akhlaghi, M. Dehghan and M. Fathi, "A New Multi-Path Routing Algorithm with Local Recovery Capability in Mobile Ad hoc Networks", Fifth International Symposium, Patras, Greece, July, 2006.
- [15] C. Sengul, R. Kravets, "Bypass Routing: an On-Demand Local Recovery Protocol for Ad Hoc Networks", Ad Hoc Networks, In press (Available online 2005).
- [16] R. Duggirala, R. Gupta, Q. Zeng, D. P. Agrawal, "Performance Enhancements of Ad Hoc Networks with Localized Route Repair", IEEE Transactions on Computers, vol. 52, no. 7, pp. 854-861, 2003.
- [17] M. khazaei, R. berangi "A multi-path routing protocol with fault tolerance in mobile ad hoc networks ", Proceedings of IEEE international CSI, 14th, tehran, iran, Oct, 2009.
- [18] UCLA Parallel Computing Laboratory and wireless Adaptive Mobility Laboratory, GloMoSim: A Scalable Simulation Environment for Wireless and Wired Network.



Mehdi khazaei received the bachelor's and master's degrees from Iran University of Science and Technology (IUST), in 2004 and 2007, respectively. Currently He is lectureship in Kermanshah University of Technology (KUT) and his researches focused on wireless networks, especially ad hoc networks.

Steganalysis of Reversible Vertical Horizontal Data Hiding Technique

Thom Ho Thi Huong,

Faculty of Information Technology,
Haiphong Private University,
Haiphong, Vietnam

Canh Ho Van

Dept. of Professional Technique,
Ministry of Public Security,
Hanoi, Vietnam

Tien Trinh Nhat

College of Technology,
Vietnam National University,
Hanoi, Vietnam

Abstract—This paper proposes a steganalysis scheme for detecting the reversible vertical horizontal (RVH) data hiding [1]. The RVH scheme was introduced in the IJCSIS International Journal Vol. 7, No. 3, March 2010. In the RVH data hiding, the message bits are embedded into cover-image by two embedding phases: the horizontal embedding procedure HEm and the vertical embedding procedure VEm. The pixel pairs belonging to the horizontally embeddable and vertically embeddable pixel pair domain are transformed to mark message bits. Through analysis, we detect out that, the two histograms of LSB scanning horizontally and vertically vary from a stego-image to the cover image. Based on this observation, we design a specific steganalytic method for attacking the RVH steganography. Experimental results show the detection accuracies of the steganography with various embedding rates are acceptable. The proposed technique can be applied in detecting the misuse of steganographic technology in malicious activities.

Keywords-Steganography, steganalysis, watermarking, cover image, stego image, payload, reversible data hiding.

I. INTRODUCTION

Steganography is [1, 3, 4, 5] is the art and science of concealed communication. The basic concept is to hide the very existence of the secret message. Digital object such as a text, image, video, or audio segment can be used as the cover data. To obtain acceptable hiding payload and keep fidelity of the stego-image, the LSB replacement techniques [2,4 or other references] are popular and widely studied in the literature. These methods usually hide more data in image areas with higher spatial variations. Reversible steganography [1,3-5] is one of the interesting branches of steganographic technology in which the original cover image can be reconstructed without any loss.

Steganalysis is the counterpart of steganography, the goal of the steganalysis is to detect the hidden message, equivalently, to discriminate the stego object from the non-stego-object. The steganalysis techniques proposed in the literature can be classified into two categories: the universal steganalysis which is designed to detect the hidden message embedded with various data embedding algorithms such as a technique proposed in [6] is used to attack the LSB steganography, and the specific steganalysis which is designed

to attack a specific steganography technique such as a steganalytic method was presented in [2] for detecting stego-images using the method proposed in [3].

In this paper, we proposed a steganalytic scheme to detect the RVH watermarking scheme introduced in brief in the abstract. Our experimental results show the feasibility of the proposed method. It is useful in detecting malicious activities on stego-images and also suggests a design consideration for future development of steganographic techniques. The rest of this paper is organized as follows. In the next section, we present again the RVH scheme in brief. Section III describes the proposed steganalytic method. Experimental results are given in section IV, and conclusions are made finally in Section V.

II. REVIEW OF THE RVH DATA HIDING SCHEME

In the steganographic method proposed in [1] used the multiple embedding strategies to improve the image quality and the embedding capacity. Basically, this method embeds each message bit b of the secret bit stream into each grayscale cover pixel pair of a grayscale cover image in raster scan order. This scheme includes two main stages, namely, the horizontal embedding procedure HEm and the vertical embedding procedure VEm. For the HEm procedure, the input image is horizontally scanned in raster scan order (i.e., from left to right and top to bottom) to gather two neighboring pixels x and y into a cover pixel pair (x, y) . If y is an odd value, then the cover pixel pair (x, y) is defined as a horizontally embeddable pixel pair. Otherwise, the cover pixel pair (x, y) is defined as a horizontally non-embeddable pixel pair. For the VEm procedure, the input image is vertically scanned in raster scan order to group two neighboring pixels u and v into a pixel pair (u, v) . If v is an even value, then the pixel pair (u, v) is defined as a vertically embeddable pixel pair. Otherwise, the pixel pair (u, v) is defined as a vertically non-embeddable pixel.

The secret bit sequence S is divided into two subsequence $S1$ and $S2$. The bit stream $B1$ is created by concatenating the secret subsequence $S1$ and the auxiliary data bit stream $A1$ (i.e., $B1=S1||A1$). Similarly, the bit stream $B2= S2||A2$. The generation of $A1$ and $A2$ will be described latter. The overview of the RVH embedding process is shown in Fig. 1.

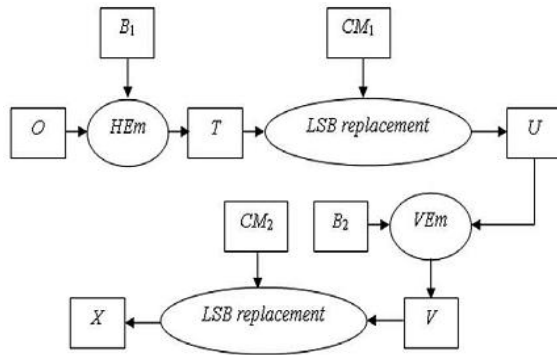


Fig. 1. Embedding phase of RVH steganographic system [1]

Firstly, the bit sequence B_1 is horizontally embedded into O by using the HEm procedure to obtain the output image T sized $H \times W$ pixels. Secondly, the compressed location map CM_1 whose length is LC_1 (will be described later), is embedded in to T by using the least significant bit (LSB) replacement technique to obtain the output image U with size of $H \times W$ pixels. Thirdly, the bit sequence B_2 is vertically embedded into U by using the VEm procedure to get the output image V with size of $H \times W$ pixels. Fourthly, the compressed location map CM_2 whose length is LC_2 is embedded into V by using the LSB replacement technique to get the final stego image with size of $H \times W$ pixels.

Each bit b in stream B_1 is horizontally embedded into each horizontally embeddable pixel pair (x, y) at a time by using the horizontal embedding rule HR defined below until the whole bit stream B_1 is completely marked into O to obtain the output image T .

Each bit b in B_2 is vertically embedded into each vertically embeddable pixel pair (u, v) at a time by using the vertical embedding rule VR defines below until the entire bit sequence B_2 is concealed into U to get the output image V .

The horizontal embedding rule HR: For each pair (x, y) , we apply the following embedding rules:

- HR1: If the to_be_embedded bit $b=1$, then the stego pixel pair is unchanged by $(x_0, y_0) = (x, y)$.
- HR2: If the to_be_embedded bit $b=0$, then the stego pixel pair is changed by $(x_0, y_0) = (x, y-1)$.

The vertical embedding rule VR: For each pair (u, v) , we apply the following embedding rules:

- VR1: If the to_be_embedded bit $b=0$, then the stego pixel pair is unchanged by $(u_0, y_0) = (u, v)$.
- VR2: If the to_be_embedded bit $b=1$, then the stego pixel pair is changed by $(u_0, y_0) = (u, v+1)$.

It is noted that the rule HR and VR don't cause the underflow and overflow problem. That is the changed pixel pairs are assured to fall in the allowable range $[0, 255]$.

The auxiliary data bit sequence A_1 is actually the LSBs of the first LC_1 (LC_1 is the length of the compressed location map CM_1 ended with the unique end of map indicator EOM_1) pixels in the image T and generated as follows. Initially, B_1 is

equal to S_1 (i.e., $B_1=S_1$). During the execution of the HEm procedure, for the first LC_1 pixels in O , when each pixel has been processed for embedding, its LSB is taken as an auxiliary data bit of A_1 and appended to the end of B_1 . That is, B_1 is gradually grown until the LC_1 auxiliary data bits in A_1 are concatenated into B_1 . Finally, the to_be_embedded bit stream is $B_1=S_1||A_1$, which is completely embedded into O .

Similar to the generation of A_1 , the auxiliary data stream A_2 is actually the LSBs of the first LC_2 (LC_2 is the length of the compressed location map CM_2 ended with the unique end of map indicator EOM_2) pixels in the image V and generated as follows. B_2 initially equals the secret bit sequence S_2 . During the execution of the procedure VEm, for the first LC_2 pixels in the image U , when each pixel has been processed for embedding, its LSB is taken as an auxiliary data bit of A_2 and append to the end of B_2 until the LC_2 auxiliary data bits in A_2 are concatenated into B_2 . Finally, the information bit sequence is $B_2=S_2||A_2$, which is fully marked into the image U .

For the purposes of extracting B_1 and recovering O , a location map HL sized $H \times (W/2)$ is needed to record the positions of the horizontally embeddable pixel pair (x, y) in O . The location map HL is a one-bit bitmap. All the entries of HL are initialized to 0. If cover pixel pair (x, y) is the horizontally embeddable pixel pair, then the corresponding entry of HL is set to be 1. Next, the location map HL is losslessly compressed by using the JBIG2 codec (Howard et al, 1998 [8]) or an arithmetic coding toolkit (Carpenter, 2002 [7]) to obtain the compressed location map CM_1 whose length is LC_1 . The compressed location map CM_1 is embedded into the image T by using the LSB replacement technique as mentioned above. Similarly, for the purposes of extracting B_2 and recovering the image U , we also need a location map VL sized $(H/2) \times W$ to mark the position of the vertically embeddable pixel pairs (u, v) in U . Then, VL is also losslessly compressed by using the JBIG2 codec or an arithmetic coding toolkit to obtain the compressed location map CM_2 whose length is LC_2 . Next, the map CM_2 is concealed into the image V by using the LSB steganography as mentioned above.

The final output of the embedding phase is the final stego image X with size of $H \times W$ pixels.

III. THE PROPOSED STEGANALYTIC SCHEME FOR THE RVH STEGANOGRAPHY

After embedding a large message sequence M (its ratio is about 90% of maximum embeddable capacity of image) into the original image Baboon sized 512×512 pixels (show Fig. 2) using the RVH scheme to obtain the stego-image Baboon, we calculate histogram of the two images (cover Baboon image and stego Baboon image), resulted in Fig. 3. It's very hard to detect any difference between the two images.

However, when we separately calculate two histograms on all pixel odd columns and all pixel even rows of the cover Baboon image, shown in Fig. 4. Similarly, calculate two histograms on all pixel odd columns and all pixel even rows of the stego Baboon image, resulted in Fig. 5. It's easy to difference between pair histogram in Fig.4 (a) and Fig. 5 (a), in Fig. 4 (b) and Fig. 5 (b). The informality appears in Fig 5 (a)

and (b) due to embedding process of RVH scheme following description in detail below.

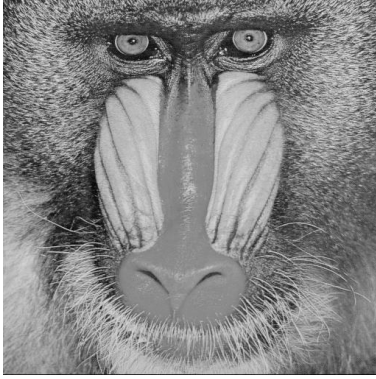


Fig. 2. The Baboon image sized 512x512 pixels

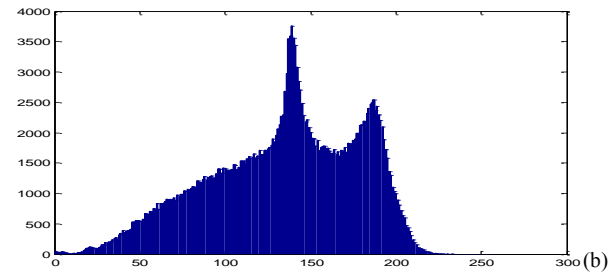
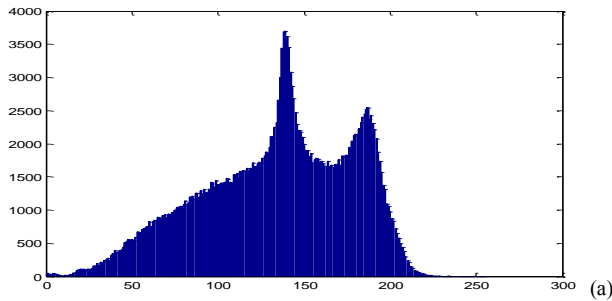


Fig. 3. Histogram of the tested two images: (a) the cover Baboon image, (b) the stego Baboon image

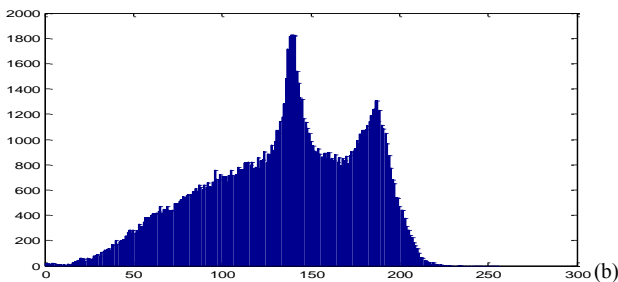
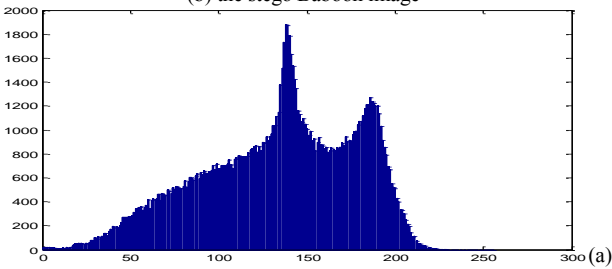


Fig. 4. Histogram of the cover Baboon images: (a) histogram on all pixel odd columns, (b) histogram on all pixel even columns

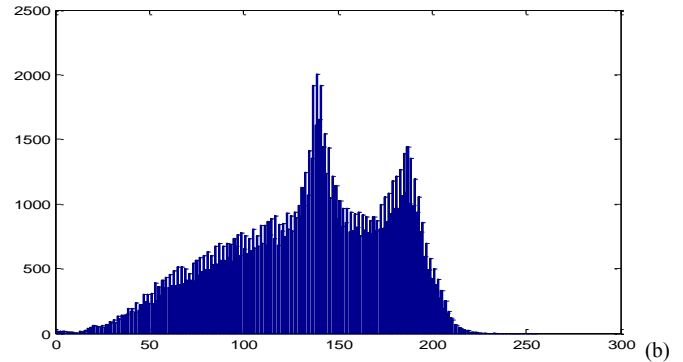
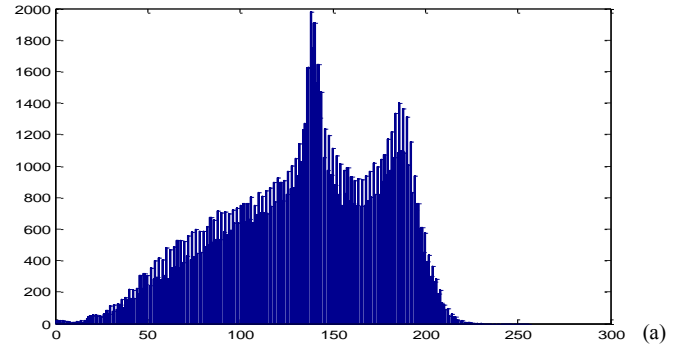


Fig. 5. Histogram of the stego Baboon images: (a) histogram on all pixel odd columns, (b) histogram on all pixel even columns

According to the horizontal embedding procedure HEm, from an input image O , the pixels of the image O are horizontally grouped into pixel pairs (x, y) , these pairs are partitioned into two sets, one set is $E1$ and other set is $\bar{E}1$, the set $E1$ contains pixels pair which are horizontally embeddable pixel pairs, while the set $\bar{E}1$ consists of those pixel pairs which are horizontally non-embeddable pixel pairs.

Now, we examine the migration of LSB histogram of the image O and the image T obtained after embedding secret bit $B1$. Without loss of generality, let (x, y) and (\tilde{x}, \tilde{y}) be the corresponding pixel pairs in the image O and the stego-image T , respectively. In the horizontal embedding procedure HEm, pixel pairs $(x, y) \in E1$, i.e. the LSB of pixel y be bit 1, are selected to embed message bits. Here, We don't examine change of LSB histogram of pixels x on pixel even-columns because they are still remained value after embedding message bits. In the image T , the LSB of \tilde{y} is changed to either 0 or 1, and each of them appears in the same probability. It is obviously that the probability of bit 0 and bit 1 are 0.5 and 0.5, respectively. For pixel pairs $(x, y) \in \bar{E}1$, i.e. the LSB of pixel y be bit 0, after embedding secret bits, \tilde{y} is unchanged. So the probability of bit 0 and 1 are 1 and 0, respectively.

Next, the compressed location map $CM1$ ($CM1$ is a binary stream, whose length is $LC1$) are marked into the image T by the LSB replacement technique to obtained image U . That changes a part of probability of LSB of bit 1 and bit 0 on all pixel even-columns in the image T . Assume that the bits are randomly distributed, so the probability of bit 0 and bit 1 are $P_{map1}(0) = P_{map1}(1)$.

Base on above discussions, the probabilities of bit 0 and bit 1 of all pixels on even-column in the image U can be calculated. Assume the probability of pixel pairs belonging to $E1$ and the probability of pixel pairs belonging to $\bar{E}1$ be P_{E1} and $P_{\bar{E}1}$, respectively. After marking the location map CM1, P_{E1} and $P_{\bar{E}1}$ are changed to P'_{E1} and $P'_{\bar{E}1}$. Let P_{R-H} is the embedding ratio defined by dividing the number pairs actually used to hide data by the total number of pairs the image O. The probability of bit $b=\{0,1\}$ of LSB of the image U can be calculated using the following equation

$$P_{LSB-H}(b) = \begin{cases} P_{R-H} \times (0.5 \times P'_{E1} + P'_{\bar{E}1}) + P_{R-\bar{H}} \times 0.5 & \text{if } b = 0 \\ P_{R-H} \times (0.5 \times P'_{E1}) + P_{R-\bar{H}} \times 0.5 & \text{if } b = 1 \end{cases} \quad (1)$$

For the vertical embedding procedure VEm, vertically scan the output image U in raster scan order to group pixel pairs (u, v) , we classify the pairs into two sets, one is $E2$ and other is $\bar{E}2$, the set $E2$ contains all pixel pairs which are vertically embeddable pixel pairs, the set $\bar{E}2$ consisting pixel pairs are vertically non-embeddable pixel pairs. Let (u, v) and (\tilde{u}, \tilde{v}) be pixel pairs of the image U (before using the procedure VEm) and the stego-image V (after embedding secret message using the procedure VEm). In the procedure VEm, only pixel pairs $(u, v) \in E2$, i.e. the LSB of v is bit 0, are embedded message bits. After embedding message bits, the LSB of the pixels \tilde{v} (obtained from $E2$) is either 0 or 1. So, the probabilities of bit 0 and bit 1 of (\tilde{u}, \tilde{v}) are equals to 0.5. For pixel pairs $(u, v) \in \bar{E}2$, i.e. the LSB of pixel v be bit 1, after embedding secret bits, \tilde{v} is unchanged. So the probability of bit 0 and 1 are 0 and 1, respectively.

Next, the compressed location map CM2 are marked into the image V by the LSB replacement technique to obtained image X. That changes a part of probability of LSB of bit 1 and bit 0 on all pixel even-rows in the image V. Assume the bits of the map CM2 are randomly distributed, so the probability of bit 0 and bit 1 are $P_{map2}(0) = P_{map2}(1)$.

From above discussions, the probabilities of bit 0 and bit 1 in the LSB of image X after using the vertical embedding procedure VEm can be calculated. Assume the probability of pixel pairs belonging to $E2$ and the probability of pixel pairs belonging to $\bar{E}2$ be P_{E2} and $P_{\bar{E}2}$, respectively. After marking the location map CM2, P_{E2} and $P_{\bar{E}2}$ are changed to P'_{E2} and $P'_{\bar{E}2}$. Let P_{R-V} be the embedding ratio defined by dividing the number pairs actually used to hide data by the total number of pairs the image V. The probability of bit $b=\{0,1\}$ of LSB of the image V can be calculated using the following equation

$$P_{LSB-V}(b) = \begin{cases} P_{R-V} \times (0.5 \times P'_{E2} + P_{R-\bar{V}} \times 0.5 & \text{if } b = 0 \\ P_{R-V} \times (0.5 \times P'_{E2} + P'_{\bar{E}2}) + P_{R-\bar{V}} \times 0.5 & \text{if } b = 1 \end{cases} \quad (2)$$

For a natural image, assume that the LSB is randomly distributed, then the expected probability of bit 0 and the probability bit 1 of all pixel on pixel even-columns are the same, i.e. $P_{LSB}(0) = P_{LSB}(1) = 0.5$. So probability $P_{E1} = 0.5$, $P_{\bar{E}1} = 0.5$. After covering a part of the LSB of image T by the location map CM1 with a probability 0.05 (assumed), the two probability P_{E1} and $P_{\bar{E}1}$ change to $P'_{E1} = 0.45$ and $P'_{\bar{E}1} = 0.55$, respectively. Consider the Baboon stego-image from the Baboon cover image, the probability of the embeddable pairs (i.e. those pixel pairs belonging to the procedure HEm) of an input image T is P'_{E1} , and 90 % of the embeddable pairs are

used to embed message, i.e. the embedding ration of $P_{R-H} = 0.45 \times 0.9 = 0.405$. From (1), we have $P_{LSB-H}(0) = 0.405 \times (0.5 \times 0.45 + 0.55) + 0.595 \times 0.5 = 0.611375$ and $P_{LSB-H}(1) = 0.405 \times (0.5 \times 0.45) + 0.595 \times 0.5 = 0.388625$. Next, calculated the probability of bit 0 and the probability of bit 1 of the output image X. We know that the probability of $E2$ equals to the probability of the LSB of bit 0 of all pixels on even-rows, i.e. $P_{E2} = P_{LSB}(0)/2 + P_{LSB-H}(0)/2 = (0.5 + 0.611375)/2 = 0.5556875$ and $P_{\bar{E}2} = 0.4443125$. After covering a part of the LSB of image V by the location map CM2 with a probability 0.05 (assumed), the two probability P_{E2} and $P_{\bar{E}2}$ change to $P'_{E2} = 0.5056875$ and $P'_{\bar{E}2} = 0.4943125$, respectively.

The embedding ratio of 90 % of the embeddable pairs are used to embed message, i.e. the embedding ration of $P_{R-V} = 0.5056875 \times 0.9 = 0.45511875$. So probability of LSB of bit 0 and bit 1 of the output image X from (2) we obtain $P_{LSB-V}(0) = 0.45511875 \times (0.5 \times 0.5056875) + 0.54488125 \times 0.5 \approx 0.3875$, $P_{LSB-V}(1) \approx 0.61248$.

Now, we check again the probability of LSB of bit 1 and the probability of LSB of bit 0 of all pixels on pixel even-columns, $P_{LSB_even_column}(0) = P_{LSB-H}(0)/2 + P_{LSB-V}(0)/2 = (0.611375 + 0.3875)/2 = 0.4994375$, $P_{LSB_even_column}(1) = (P_{LSB-H}(1) + P_{LSB-V}(1))/2 = (0.388625 + 0.61248)/2 = 0.5005525$. We found out that the probability of bit 0 $P_{LSB_even_column}(0)$ and probability of bit 1 $P_{LSB_even_column}(1)$ are the same, that is after completing the vertical procedure VEm, it make the value of these probabilities be balanced. However, the probability of LSB of bit 0 and bit 1 of all pixels on pixel odd-columns don't equal based on the following calculating: $P_{LSB_odd_column}(0) = (P_{LSB_org_odd_column}(0)/2 + P_{LSB-V}(0)/2) = (0.5/2 + 0.3875/2) = 0.44375$, $P_{LSB_odd_column}(1) = (P_{LSB_org_odd_column}(1)/2 + P_{LSB-V}(1)/2) = (0.5/2 + 0.61248/2) = 0.55624$. Where $P_{LSB_org_odd_column}(0)$ and $P_{LSB_org_odd_column}(1)$ be the probabilities of the LSB of bit 0 and bit 1 of all pixels on the pixel odd-column of the image X. A half of them isn't changed during process of the RVH scheme, so $P_{LSB_org_odd_column}(0)/2$ and $P_{LSB_org_odd_column}(1)/2$ equal to 0.5/2 and 0.5/2, respectively. We can see obvious difference of the occurrences of bit 1 and bit 0 in the LSB on all pixel odd-columns and all pixel even-column of the stego-image of the RVH scheme with respect to a standard natural image. Based on the problem, the following rule is given to discriminate a stego-image of the RVH steganography from a nature image.

$$W(X) = \begin{cases} \text{true, if } |P_{LSB}(0) - P_{LSB}(1)| > T \\ \text{false, otherwise} \end{cases} \quad (3)$$

From equation (3), an image is detected be stego-image marked by the RVH scheme if one of the measured values $|P_{LSB}(0) - P_{LSB}(1)|$ on all pixel odd columns (or pixel even columns) or pixel even rows (or pixel odd rows) is greater than threshold T ($0 \leq T \leq 1$). The threshold T is used to control the decision boundary of nature images and stego images, its value depends on specific applications.

IV. EXPERIMENTAL RESULT

To show the reliability of the proposed method, we take 500 image from USC-SIPI Image Database [9] and content based image retrieval (CBIR) image database [10] and convert them into 8-bit grayscale images. The images are used to test

proposed classification. The hidden messages used in our test are made by the pseudo random number generator. We embed different amount of message using the RVH scheme, and measure the migration of LSB histogram in the stego images. Five embedding ratios 0%, 25%, 50%, 75% and 100% are used in the test, and the obtained $|P_{LSB(0)} - P_{LSB(1)}|$ values on all pixel even-rows of the stego-images are depicted in Fig. 6 – 10, respectively. We also measure the accuracy of the proposed method in detecting the RVH scheme in different embedding ratio and likelihood threshold value T , shown in table 1.

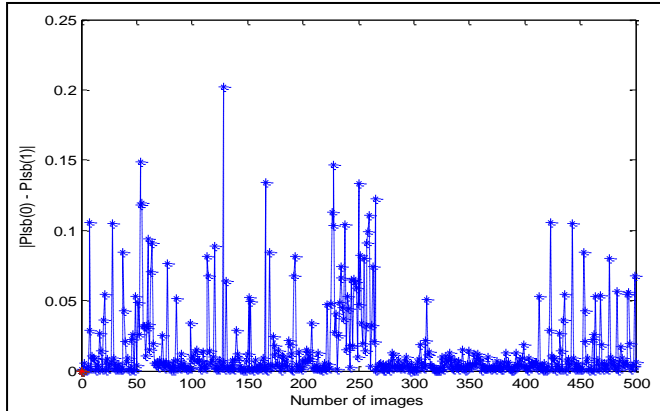


Fig. 6. The distribution of $|P_{LSB(0)} - P_{LSB(1)}|$ value of the 500 cover images on all pixel even-rows

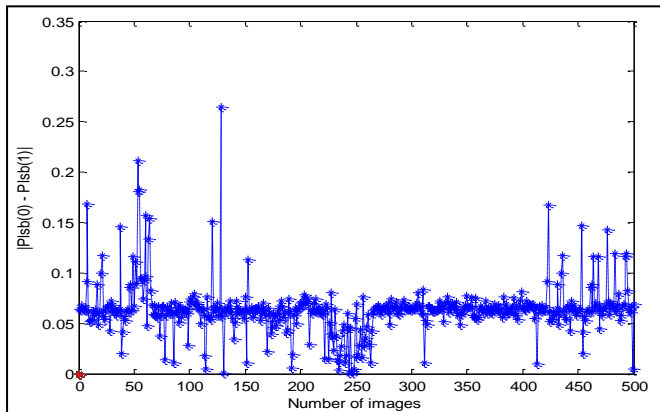


Fig. 7. The distribution of $|P_{LSB(0)} - P_{LSB(1)}|$ value of the 500 stego images on all pixel even-rows with embedding ratio 25%

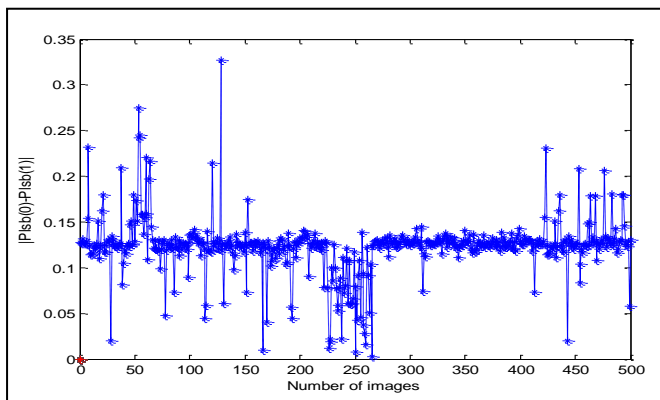


Fig. 8. The distribution of $|P_{LSB(0)} - P_{LSB(1)}|$ value of the 500 stego images on all pixel even-rows with embedding ratio 50%

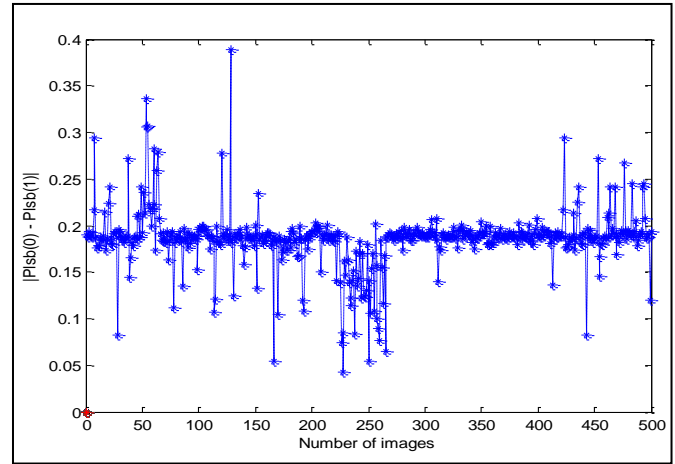


Fig. 9. The distribution of $|P_{LSB(0)} - P_{LSB(1)}|$ value of the 500 stego images on all pixel even-rows with embedding ratio 75%

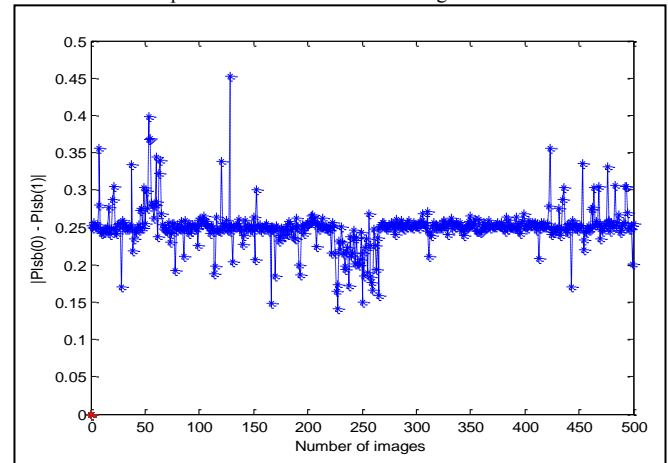


Fig. 10. The distribution of $|P_{LSB(0)} - P_{LSB(1)}|$ value of the 500 stego images on all pixel even-rows with embedding ratio 100%

TABLE I. THE DETECTION ACCURACY OF THE PROPOSED METHOD WITH VARIOUS EMBEDDING RATIOS AND THRESHOLD VALUES

Embedding Ratio (%)		0	25	50	75	100
Threshold T						
0.01	Cover	70.6 %	2%	0.6%	0 %	0 %
	Stego	29.4 %	98%	99.4 %	100%	100 %
0.02	Cover	80.6 %	5.8 %	0.6 %	0%	0
	Stego	19.4 %	94.2 %	99.4 %	100%	100 %
0.03	Cover	84.4 %	7.4 %	2.2 %	0	0
	Stego	15.6 %	92.6 %	97.8 %	100%	100%
0.04	Cover	86.8 %	11 %	2.4 %	0	0
	Stego	13.2 %	89 %	97.6 %	100%	100%
0.05	Cover	89 %	13.8 %	3.6 %	0.2 %	0
	Stego	11 %	86.2%	96.4 %	99.8%	100%

From Fig. 6 , we can see that most of the value of $|P_{LSB(0)} - P_{LSB(1)}|$ approach zero for natural images, while the higher value of $|P_{LSB(0)} - P_{LSB(1)}|$ is obtained with embedding ratio 25%, 50%, 75% and 100% shown from Fig. 7 to Fig. 10. From table 1, we see that when the likelihood threshold value T is set

0.035, we can obtain an acceptable result in detecting stego images used the RVH steganography.

V. CONCLUSION

The paper presents a method to break the RVH steganography based on the observation of the distribution of 0 and 1 bits of the LSBs on pixel odd-columns (pixel even-columns) or pixel even-rows (pixel odd-rows) of the RVH stego-images. The experimental results are shown that the proposed method can detect stego - images reliably with embedding ratio being greater 25%. On the other hand, we show a problem of security of the RVH scheme in the data embedding process.

ACKNOWLEDGMENT

Our special thanks to Haiphong Private University (HPU) for their financial support to our research and College of Technology, Vietnam National University, Hanoi for their support to good working environment. We would like to extend our thanks to my guide, our friends and family members without whose inspiration and support our efforts would not have come to success.

REFERENCES

- [1] P. Mohan Kumar, K. L. Shunmuganathan, *A reversible high embedding capacity data hiding technique for hiding secret data in images*, International Journal of Computer Science and Information Security, Vol.7, No. 3, March 2010, pp. 109-115.
- [2] Yeh-Shun Chen, Ran-Zan Wang, Yeuan-Kuen Lee, Shih-Yu Huang, *Steganalysis of reversible contrast mapping water marking*, Proceedings of the world congress on Engineering 2008 Vol I, WCE2008, July 2-4, 2008, London, U.K., pp. 555-557.
- [3] D. Coltuc and J. M. Chassery, "Very fast watermarking by reversible contrast mapping," IEEE Signal Processing Lett., vol. 14, no. 4, pp. 255-258, Apr. 2007.
- [4] J. Tian, "Reversible Data embedding using a difference expansion," IEEE Trans. Circuits Syst. Video technol., vol. 13, no. 8, pp. 890-896, Aug. 2003.
- [5] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible Data Hiding," IEEE Trans. Circuits Syst. Video technol., vol. 16, no. 3, pp. 354-362, Mar. 2006.
- [6] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," Proceedings of the ACM International Multimedia Conference and Exhibition, pp. 27-30, 2001.

- [7] Carpenter, B., 2002. Compression via Arithmetic Coding <http://www.colloquial.com/ArithmeticCoding/>
- [8] P.G. Howard, F. Kossentini, B. Martins, S. Forchhammer, W. J. Rucklidge, 1998. The emerging JBIG2 standard. IEEE Transactions on Circuits and Systems for Video Technology 8 (7), pp. 838-848.
- [9] USC-SIPI Image Database, "http://sipi.usc.edu/services/database/Database.htm"
- [10] CBIR Image Database, University of Washington, [http://www.cs.washington.edu/research/imagetatabase/groundtruth/..](http://www.cs.washington.edu/research/imagetatabase/groundtruth/)

AUTHORS PROFILE



Ho Thi Huong Thom received the B.S. degree of Information Technology department from Haiphong Private University and the M.S. degree in Information Systems from College of Technology, Vietnam National University in Vietnam, in 2001 and 2005, respectively.

She has started her career as Lecturer in Department of Information Technology in Haiphong Private University, Vietnam and served for 9 years. Currently, she is pursuing Doctor of Information Systems from College of Technology, Vietnam National University, Hanoi, Vietnam. Her research interests includes Image processing, Information Security, Information Hiding.



Ho Van Canh received the B.S. degree in Mathematics from Hanoi City University in Vietnam in 1973, the Dr. Sci. degree in Faculty of statistology from KOMENSKY University in Czechoslovakia in 1987. Currently, he has been working as a cryptologist in Dept. of Professional Technique, Ministry of Public Security, Vietnam. His

research interests include cryptography, information security, information hiding.



Trinh Nhat Tien received the B.S degree from University of Prague in Czechoslovakia in 1974, and the Dr. degree from University of Prague, Czechoslovakia and University of Hanoi, Vietnam in 1984. He has started as Lecturer in Department of Information Technology of College of

Technology, Vietnam National University, Hanoi, Vietnam since 1974. His research interests include algorithm, complexity of algorithm, information security.

Off-line Handwritten Signature Recognition Using Wavelet Neural Network

Mayada Tarek¹

Computer Science Department,
Faculty of Computers and Information
Sciences,
Mansoura, Egypt

Taher Hamza

Computer Science Department,
Faculty of Computers and Information
Sciences,
Mansoura, Egypt

Elsayed Radwan

Computer Science Department,
Faculty of Computers and Information
Sciences,
Mansoura, Egypt

Abstract—Automatic signature verification is a well-established and an active area for research with numerous applications such as bank check verification, ATM access, etc. Most off-Line signature verification systems depend on pixels intensity in feature extraction process which is sensitive to noise and any scale or rotation process on signature image. This paper proposes an off-line handwritten signature recognition system using Discrete Wavelet Transform as feature extraction technique to extract wavelet energy values from signature image without any dependency of image pixels intensity. Since Discrete Wavelet Transform suffers from down-sample process, Wavelet Neural Network is used as a classifier to solve this problem. A comparative study will be illustrated between the proposed combination system and pervious off-line handwritten signature recognition systems. Conclusions will be appeared and future work is proposed.

Keywords—Discrete Wavelet Transform (DWT); Wavelet Energy; Wavelet Neural Network (WNN); Off-line Handwritten Signature.

I. INTRODUCTION

In the field of personal identification, two types of biometrics means can be considered; first, physiological biometrics, which involves data derived from the direct measurement of some part of the human body; for-example fingerprint-, face-, palm print-, retina-based verification. Second, behavioural biometrics, which involves data derived from an action taken by a person, or indirectly measures characteristics of the human body; for-example: speech-, keystroke dynamics and signature-based verification [1].

In the last few decades, researchers have made great efforts on off-line signature verification [1] for-example; using the statistics of high grey-level pixels to identify pseudo-dynamic characteristics of signatures; developing technique based on global and grid features

in conjunction with a simple Euclidean distance classifier; proposing a system for off-line signature verification consists of four subsystems based on geometric features, moment representations, envelope characteristics and wavelet features; applying wavelet on signature verification [2,3,4,5].

Although these methods achieved a good results, they still suffer from the exchangeability of signature rotation and the distinguish-ability of person signature size. Most of these feature extraction methods depend on signature shape or pixels intensity in specific region of signature. However, pixels' intensity are sensitive to noise and also the signature shape may vary according to translation, rotation and scale variations of signature image [6].

Two types of feature can be extracted from signature image; first, global features which are extracted from the whole signature, including block codes [7]; second, local features which are calculated to describe the geometrical and topological characteristics of local segments [8]. Because of the absence of dynamic information in offline verification system, global features extraction are most appropriate [9]. One of the most appropriate global features extraction techniques is wavelet transform, since it extracts time-frequency wavelet coefficients from the signature image [8]. Wavelet Transform is especially suitable for processing an off-line signature image where most details could be hardly represented by functions, but could be matched by the various versions of the mother wavelet with various translations and dilations [10]. Also, wavelet transform is invariant to translation, rotation and scale of the image. Because of the advantage of wavelet transform, this paper uses it in feature extraction stage.

Since one of problems that face wavelet is the huge size of its coefficients, statistical model can be introduced to represent them. This paper uses wavelet energy as statistical model to represent all wavelet coefficients in efficient way. Another problem is down-sample process which can lose some important extracted feature from signature image[11]. This paper proposes a Wavelet Neural Network (WNN) technique for off-line signature recognition to overcome the disadvantages of Discrete Wavelet Transform (DWT) down-sample process.

¹Corresponding Author

Mail: mayaatarek@yahoo.com

Tel : 020108631688

WNN takes full advantages of the partial-resolution characteristic of the wavelet transform and the nonlinear mapping behaviour of Artificial Neural Networks (ANN) [15].

This paper proposes a combination model between DWT and WNN techniques for off-line handwritten signature recognition system. DWT technique will analysis signature image to extract wavelet detail coefficients. To reduce the huge number of these coefficients with the same accuracy, a statistical model is represented by wavelet energy. Because of the problem of down sample, WNN technique will be used as a suitable classifier technique to overcome this problem. Also, a modified back-propagation technique is used in learning WNN. A testing stage examines the unseen signature. Moreover, a comparative study will be illustrated between the proposed combination system and pervious off-line handwritten signature recognition systems. Conclusions will be appeared and future work is suggested.

The rest of this paper organized as; in Section 2, Handwritten signature, wavelet transform (WT), Wavelet Neural Network (WNN) are mentioned. Methodology and applications using a combination between DWT and WNN techniques is described in Section 3. Section4, consists of the result of the proposed combination system and a comparative study between three strategies (signature image pixels intensity value as input to ANN, signature wavelet energy values as input to ANN and signature wavelet energy values as input to WNN). Finally section 5 concludes the paper.

II. PRELIMINARIES

A. Handwritten Signature

Handwritten signatures are widely accepted as a means of document authentication, authorization and personal verification. For legality most documents like bank cheques, travel passports and academic certificates need to have authorized handwritten signatures. In modern society where fraud is rampant, there is the need for an automatic Handwritten Signature Verification system (HSV) [6]. Dependency on automation is due to the difficulty faced in visual assessment for different types and different sizes of signatures. Simple, cursive, graphical and not a connected curve pattern are some of the different types of signatures and machines are far superior when it comes to processing speed and management of large data sets with consistency [12].

Automatic HSV systems are classified into two types: offline HSV and online HSV: static or off-line system and dynamic or on-line system. Static off-line system gain data after writing process has been completed. In this case the signature is represented as a grey level

image. Dynamic systems use on-line acquisition devices that generate electronic signals representative of the signature during the writing process [1].

It is well known that no two genuine signatures of a person are precisely the same and some signature experts note that if two signatures written on paper were same, then they could be considered as forgery by tracing. Unfortunately, off-line signature verification is a difficult discrimination problem because of dynamic information regarding the signing velocity, pressure and stroke order are not available also an off-line handwritten signature is depend for instance on, the angle at which people sign may be different due to seating position or due to support taken by hand on the writing surface and all this information can't be extract from static image[12].

B. Wavelet Transform :

Wavelet Transform (WT) [13] is become a powerful alternative analysis tool to Fourier methods in many signal processing applications. The main advantages of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones, thus leading to an optimal time-frequency resolution in all the frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets lack the requirement of stationary. There are two types of Wavelet Transform; Continous Wavelet Transform(CWT), Discrete Wavelet Transform (DWT).

The Continuous Wavelet Transform [14] of a 1-D signal $x(t)$ is defined as in *equation (1)*:

$$\omega_{(a,b)}(t) = \frac{1}{\sqrt{|a|}} \int_t x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

Where $\psi(t)$ is the mother wavelet or the basis function which, in a form analogous to sines and cosines in Fourier analysis. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) b and scaling (dilation or compression) a .

The Discrete Wavelet Transform [14], which is based on sub-band coding is found to yield a fast computation of wavelet transform. It is easy to implement and reduces the computation time and resources required.

In CWT, the signals are analyzed using a set of basis functions which relate to each other by simple scaling and translation. In the case of DWT, a time-scale representation of the digital signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut off frequencies at different scales[14].

In DWT, the extension to 2-D is usually performed by using a product of 1-D filters. The transform is computed by applying a filter bank as shown in *Figure 1*. L and H to denote the 1-D low pass and high

pass filter, respectively. The rows and columns of image are processed separately and down sampled by a factor of 2 in each direction which may cause losing important feature. Resulting in one low pass image LL and three detail images HL, LH, and HH. **Figure 2a** shows the one-level decomposition of **Figure 1** in the spatial domain. The LH channel contains image information of low horizontal frequency and high vertical frequency, the HL channel contains high horizontal frequency and low vertical frequency, and the HH channel contains high horizontal and high vertical frequencies. Three-level frequency decomposition is shown in **Figure 2b**. Note that in multi-scale wavelet decomposition only the LL sub-band is successively decomposed [13].

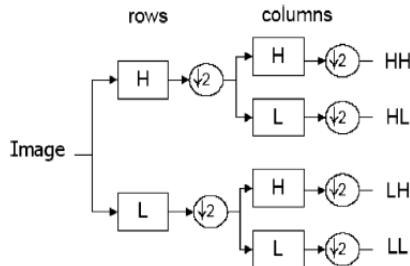
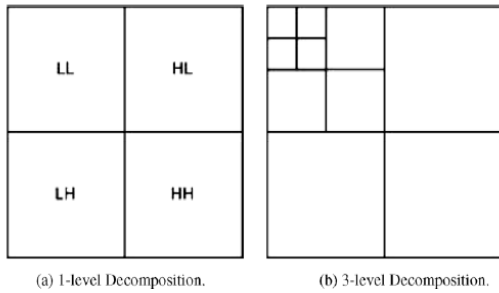


Figure 1: A one-level wavelet analysis filter bank.



(a) 1-level Decomposition. (b) 3-level Decomposition.
Figure 2 : Wavelet frequency decomposition.

C. Wavelet Neural Network :

WNN is a combination technique between neural network and wavelet decomposition. The advantages of the WNN are a high-speed learning and a good convergence to the global minimum [15]. The reason for the application of WNN in case of such a problem as classification is that the feature extraction and representation properties of the wavelet transform are merged into the structure of the ANN to further extend the ability to approximate complicated patterns [16].

The WNN can be considered an expanded perceptron [17]. The WNN is designed as a three-layer structure with an input layer, a wavelet layer, and an output layer. The topological structure of the WNN is illustrated in **Figure 3**.

In WNN, both the position and dilation of the wavelets as well as the weights are optimized. The basic neuron of a WNN is a multidimensional wavelet in which the dilation and translation coefficients are considered as

neuron parameters. The output of WNN is therefore a linear combination of several multidimensional wavelets [15].

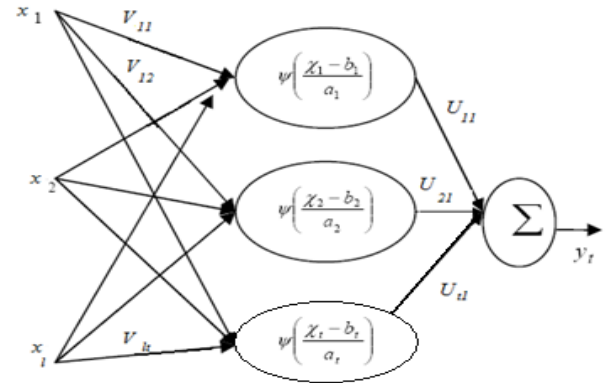


Figure 3 : The structure of the Wavelet Neural Network

In this WNN model, the hidden neurons have wavelet activation functions ψ and have two parameter a_i, b_i which represent dilation and translation parameter of wavelet function and V is the weight connecting the input layer and hidden layer and U is the weight connecting the hidden layer and output layer.

Let $X^n = \{x_i\}, i=1, \dots, L$ and $n=1, \dots, N$ be the WNN input to no. n sample ; $Y^n = \{y_k\}, k=1, \dots, S$ represents the output of WNN ; $D = \{d_k\}, k=1, \dots, S$ represents the expected output ; V_{ij} represents the connection weight between no. i node (input layer) and j node (hidden layer) ; U_{jk} represents the connection weight between no. j node (hidden layer) and k node (output layer) . Where N is the number of Sample ; S is the number of output node ; L is the number of input node ; M is the number of hidden layer.

III. WAVELET NEURAL NETWORK FOR OFF-LINE HANDWRITTEN SIGNATURE RECOGNITION

According to the fact that there aren't two genuine signatures of one person are precisely the same, many efforts have been done in order to comprehend the delicate nuances of person signatures [12]. Especially off-line signature recognition needs more effort because of the absence of dynamic information that can't be extracted from static image [12]. Also, the problems of translation, rotation and scale variation of signature image are still found when dealing with signature image pixels' intensity [6].

This paper presents an implementation for off-line handwritten signature recognition system using DWT technique in feature extraction phase and WNN in classification phase to overcome all the above problems with off-line handwritten signature recognition system. DWT technique depends on analyzing all signature shapes (continuous case) instead of analyzing the pixels intensity or segmentation part of signature (discrete case). Because of the problem of down-sample caused

by DWT technique, WNN technique will be used in classification stage to overcome this problem.

The proposed Off-line Handwritten Signature recognition system as depicted in **Figure 4** involves four stages:

- ❖ Scan and removing noise stage.
- ❖ Feature extraction stage.
- ❖ Classification stage.
- ❖ Test stage.

First stage, **Scan and removing noise stage**, each off-line handwritten signature is scanned due to creating signature image. Because of the scanning process, removing noise from signature image is an important task. In this paper, the median filter [18] is used to remove noise for two reasons. First, it preserves the structural shape of the signature without removing small strokes. Second, the absence of dealing with median filter in wavelet transform technique, which work to analysis image with low/high-pass filters corresponding to its wavelet function.

The median filter is a nonlinear digital filtering technique which is often used to remove noise. Noise reduction is a typical pre-processing step that improves the results. The median filter considers each pixel in the image in turn and looks at its nearby neighbours to decide whether or not the pixel intensity value is representative of its surroundings. The median filter replaces the pixel with the *median* of its neighbouring pixel intensity values. The median is calculated by first sorting all the pixel intensity values from the surrounding neighbourhood into numerical order and then replacing the pixel being considered with the middle pixel intensity value [19].

Second stage, **Feature extraction stage** is the most important component for designing the intelligent system based on pattern recognition. The pattern space is usually of high dimensionality. The objective of the feature extraction is to characterize the object by reducing the dimensionality of the measurement space (i.e., the original waveform). The best classifier will perform poorly if the features are not chosen well [20].

According to the fact that there aren't two genuine signatures of one person are precisely the same, the differences in the same person signature may exist in details. Because of the details of an image will access by high pass filter, DWT is used to access high pass information of person's signature images. This information is fused to obtain pattern of each person's signatures that contains all details information of his/her signatures [21]. Details information extracted by DWT technique must be extracted using suitable wavelet function to off-line handwritten signature recognition application. According to the previous work in off-line handwritten signature recognition have apply Daubechies 4, 12 and 20 wavelets functions as depicted in **Figure 5** [5] as a mother wavelet function, which can preserve maximum details of the original image, reflect outline of the image objectively and decrease the FRR.

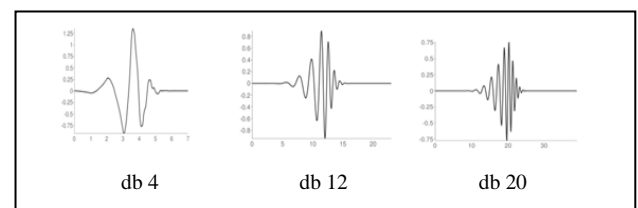


Figure 5: Daubechies 4, 12 and 20 wavelets functions

After DWT is applied on the image, wavelet coefficients from the approximation sub-band is discard and interested in wavelet coefficients from the details sub-bands of all the decomposition levels . This entire coefficient is very large to be used as feature extraction model from an image. These wavelet coefficients can be represented as statistical features such as mean, median, standard deviation, energy and entropy [22]. In this paper, wavelet energy values for details wavelet sub-band is the reduced vector that contain the main information that represent person signature from the huge wavelet decomposition values.

While off-line handwritten signature image is sensitive to translation, rotation and scale changes; the same images with different scale or rotational may have different wavelet coefficients. The main reason is that the efficient implementation of 2D-DWT requires applying a filter bank along the rows and columns of an image [23].

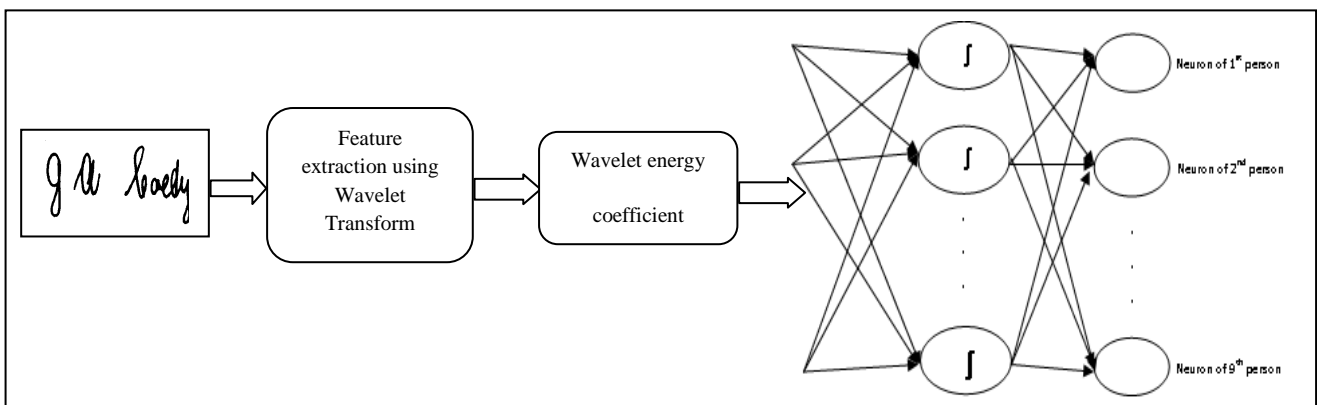


Figure 4: Proposed off-line Handwritten signature Recognition System

Due to the separability of the filters, the separable 2D-DWT is strongly oriented in the horizontal and vertical directions. This makes it hardly possible to extract translation, rotation and scale invariant features from the wavelet coefficients. Wavelet energy can keep the main characteristic of these wavelet coefficients and make the same images with different translation, rotation and scale having the same wavelet energy values[23]. Wavelet energy values can be computed after analysis signature image to it's wavelet sub-image coefficient at three level analysis (LLx, HLx, LHx, HHx). The percentages of energy of these high frequency sub-images at the k -level wavelet decomposition is defined in *equation (2,3,4)*[24]:

$$EHL^{(K)} = \frac{100 * \sum(HL \text{ decomposition vector at level } K)^2}{\sum(\text{decomposition vector})^2} \quad (2)$$

$$ELH^{(K)} = \frac{100 * \sum(LH \text{ decomposition vector at level } K)^2}{\sum(\text{decomposition vector})^2} \quad (3)$$

$$EHH^{(K)} = \frac{100 * \sum(HH \text{ decomposition vector at level } K)^2}{\sum(\text{decomposition vector})^2} \quad (4)$$

Third stage, **Classification stage**, after we get the suitable wavelet energy values that represent signature image, we take this values as input to WNN and train this network with a modified Back-propagation (BP) training algorithm to get efficient off-line signature recognition. Using WNN for two reasons; first, traditional ANN has many trade-off because of complex computations, huge iterations and learning algorithms are responsible for slowing down the recognition rate using ANN; second ,recover losing important information from signature image in DWT technique because of down-sample process as depicted in *Figure1*.

The back-propagation algorithm seems to be superior in this handwritten signature verification environment [25]. In a back-propagation neural network[26], the learning algorithm has two phases. First, a training input pattern is presented to WNN input layer. The WNN propagates the input pattern from layer to layer until the output pattern is generated by the output layer. If this pattern is different from the desired output, an error is calculated and then propagated backward through the WNN from the output layer to the input layer. The weights and both the position and dilation of the wavelets layer are modified as the error is propagated. The modified back-propagation training algorithm in WNN [27]as shown in *Figure 6*.

In this work, The input layer represents wavelet energy values feature vector to neural network. The output layer represents the ability to recognize the human signature. The middle layer determined the ability to learn the person signature recognition. Because of the ability of Morlet function to deal with big input domain [28] and represents its wave form in equation, Morlet function will be the suitable wavelet activation functions ψ in WNN to recognize offline handwritten signature application. Morlet function equation and it's derivation in *equation (5,6)*[27]:

$$\psi(x) = \cos(1.75x) \exp\left(-\frac{x^2}{2}\right) \quad (5)$$

Then.

$$\frac{\partial \psi(x)}{\partial x} = -[x \cos(1.75x) + 1.75 \sin(1.75x)] \exp\left(-\frac{x^2}{2}\right) \quad (6)$$

Input : wavelet energy values extracted from signature image

Output : Class of recognized Signature

➤ **Step1: Initialize weights and offsets.**

Set all weights and node offsets to small random values.

Initialize position and dilation parameter for each wavelet neuron in wavelet layer.

To choose centre point (p) between interval $[z_1, z_2]$ (input domain), then

$$b_1 = p, \quad a_1 = 0.5(z_1 - z_2)$$

Interval $[z_1, z_2]$ is divided into two parts by point p.

In each sub-interval, we recursively repeat the same procedure which will

initialize b_2, a_2 and b_3, a_3 and so on, until all the wavelet are initialize.

➤ **Step2: Present input and desired outputs**

Present a continuous valued input vector X_1, X_2, \dots, X_L

and specify the desired output D_1, D_2, \dots, D_S .

If the net is used as a classifier then all desired outputs are typically set to zero except for that corresponding to the class the input is from. That desired output is 1. The input could be new on each trial or samples from a training set could be presented cyclically until stabilize.

➤ **Step 3: Calculate Actual Output**

$$Y_k^n = \sum_{j=1}^M U_{kj} \psi \left[\frac{\sum_{i=1}^L V_{ij} X_i^n - b_j}{a_j} \right] \quad (7)$$

➤ **Step 4: Calculate Error function**

$$E = \frac{1}{2N} \sum_{n=1}^N \sum_{k=1}^S (Y_k^n - D_k^n)^2 \quad (8)$$

➤ **Step 5: Propagate error to weights and position and dilation parameter**

$$\frac{\partial E}{\partial U_{jk}} = \frac{1}{N} \sum_{n=1}^N (Y_k^n - D_k^n) \psi \left[\frac{\sum_{i=1}^L V_{ij} X_i^n - b_j}{a_j} \right] \quad (9)$$

$$\frac{\partial E}{\partial V_{ij}} = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^S [(Y_k^n - D_k^n) U_{jk} \frac{\partial \psi(T)}{\partial T} \frac{X_i^n}{a_j}] \quad (10)$$

Where $T = \frac{\sum_{i=1}^L V_{ij} X_i^n - b_j}{a_j}$

$$\frac{\partial E}{\partial a_j} = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^S (Y_k^n - D_k^n) U_{jk} \frac{\partial \psi(T)}{\partial T} \left[-\frac{\sum_{i=1}^L V_{ij} X_i^n - b_j}{a_j^2} \right] \quad (11)$$

$$\frac{\partial E}{\partial b_j} = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^S (Y_k^n - D_k^n) U_{jk} \frac{\partial \psi(T)}{\partial T} \left(-\frac{1}{a_j} \right) \quad (12)$$

➤ **Step 6: Update weights and position and dilation parameter**

$$U_{jk}^{ii+1} = U_{jk}^{ii} - \alpha \frac{\partial E}{\partial U_{jk}} - \varepsilon (U_{jk}^{ii} - U_{jk}^{ii-1}) \quad (13)$$

where α is learning rate

ε is momentum factor

➤ **Step 7: Repeat by going to step 2**

Figure 6: Back-propagation training algorithm in Wavelet Neural Network

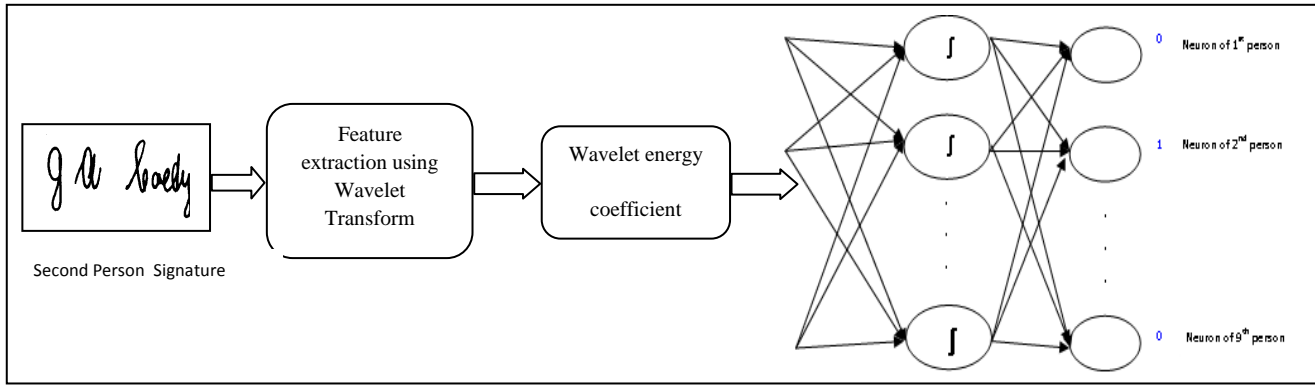


Figure 7: Testing stage of off-line Handwritten signature Recognition System

Finally, **Test stage**, after learning WNN, we can examine the ability of WNN to verify the signature of any person as shown in **Figure 7**. In this stage our goal is to input signature image and recognize the person signature. After scanning and removing noise from person signature image, wavelet coefficients produce after analysis image with DWT technique and then compute wavelet energy value from wavelet detail coefficients, finally, this wavelet energy values are taken as input to test WNN classifier to find result in output layer with only 1 value in only one neuron. Number of neurons in the output layer represents the number of person that system recognize.

IV. RESULT :

This section summarizes the results of using DWT technique (wavelet energy values) as feature extraction technique and WNN as classifier to off-line handwritten signature recognition system. This paper uses nine person handwritten signatures as show in **Figure 8**, each person has twenty image of his handwritten signature ,ten for train stage and ten for test stage .

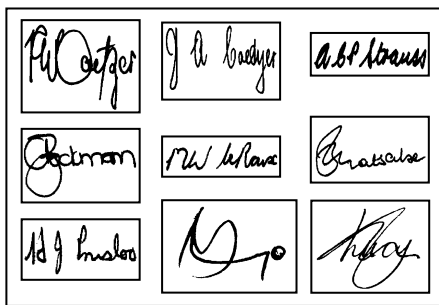


Figure 8: Sample Signature images

In **feature extraction stage**, wavelet detail coefficients are extracted from signature image using (db4 or db 12 or db20) wavelet function. To determine the suitable wavelet function to our database, WNN is used as a classifier to evaluate the suitable one. Wavelet detail coefficients (at one level analysis) of signature image according to one wavelet function is taken as trained data to WNN.

Three WNN will be found to compare the recognition rate between tree wavelet function. Modified BP training algorithm as in **Figure 6** is used to train WNN. Finally, testing WNN with trained signature . **Figure 9** shows the recognition rate to (db4,db12,db20) wavelet detail coefficients using WNN as mention above. As a result from **Figure 9**, Db20 is recognizing to be the suitable wavelet function which have high recognition rate in our database to offline handwritten application.

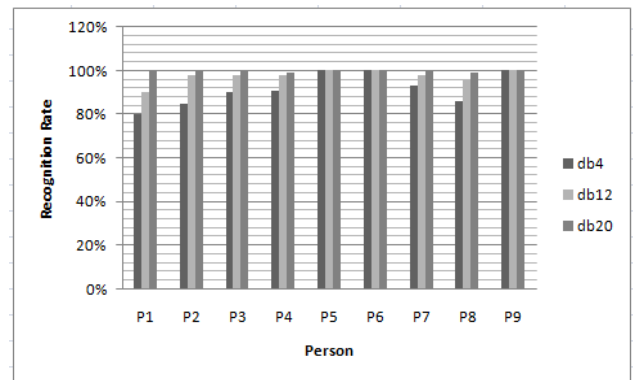


Figure 9: Offline handwritten signature recognition rate using (db4,db12,db20) wavelet detail coefficients

After determine the suitable extracting wavelet function, wavelet energy from each signature image is computed using **equation 2,3,4** with db20 as wavelet function at three level analysis. Nine wavelet energy coefficients are represented each signature image.

Table 1: WNN architecture and training parameters

The number of layers	3
The number of neuron on the layers	Input:9 Hidden:18 Output:9
The initial weights and biases	Random
Wavelet Activation functions	Morlet function
Learning rule	Back-Propagation
MSE	0.0001
Learning rate	0.1
Momentum factor	0.009

In **classification stage**, WNN is used with parameters shown in **Table1**. These parameters are selected for WNN structure after several different experiments. In these experiments, the WNN is employed with different parameters such as the number of hidden layers, the size of the hidden layers, value of the moment constant and learning rate, and type of the activation functions. Wavelet energy values for each signature image are the input features to WNN input layer. Each neuron in WNN output layer represent a person.

In the **test stage**, Applying the test wavelet energy values of the test signature to trained WNN. Evaluating the proposed off-line handwritten signature recognition system by recognition rate to each person as shown in **Figure 10**.

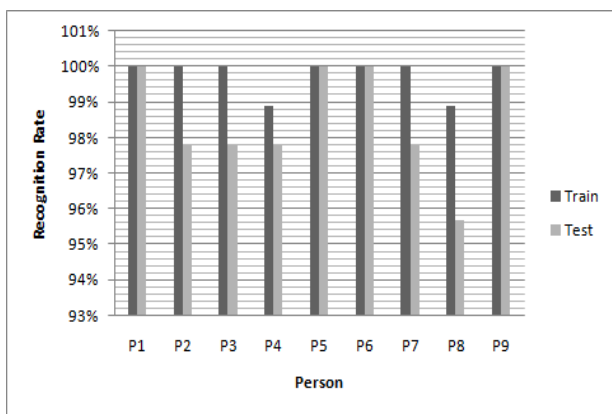


Figure 10: Proposed off-line handwritten signature recognition system result

All system evaluation is made by two concept False Acceptance Rate(FAR) which indicates how many forgeries were incorrectly classified as genuine signatures ,and False Rejected Rate(FRR) which indicates how many genuine signatures were incorrectly rejected by the system. To the training signatures FAR and FRR is 0.01% and to the testing signatures FAR and FRR is 0.07% .

To evaluate our proposed system a comparative study between three off-line handwritten signature systems is made:

- 1-signature image pixels intensity value as input to ANN(ANN)
- 2- signature wavelet energy values as input to ANN (WE+ANN)
- 3- Our proposed system signature wavelet energy values as input to WNN (WE+WNN).

Figure 11 represent the recognition rate to each training person data and **Figure 12** represent the recognition rate to each testing person data. **Figure 11** and **Figure 12** concluded that our proposed system has the highest recognition rate.

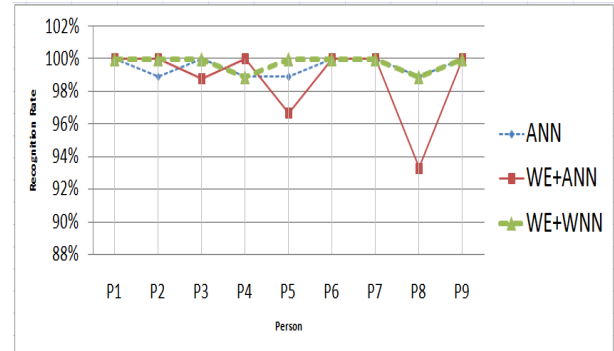


Figure 11: Comparative Study between signature image pixels intensity value as input to ANN (ANN)and signature wavelet energy values as input to ANN (WE+ANN)and signature wavelet energy values as input to WNN (WE+WNN)with training data.

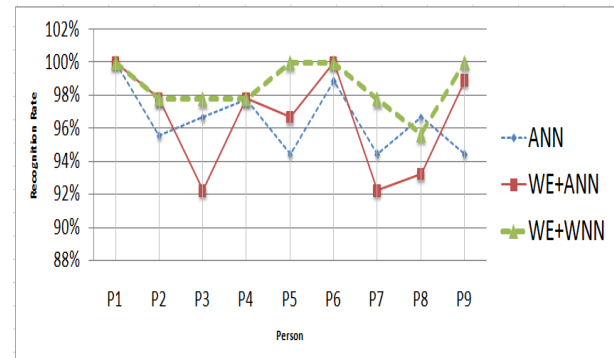


Figure 12: Comparative Study between signature image pixels intensity value as input to ANN (ANN)and signature wavelet energy values as input to ANN (WE+ANN)and signature wavelet energy values as input to WNN (WE+WNN) testing data.

V. CONCLUSIONS AND FUTURE WORK

Handwritten signature recognition plays an important role in our daily life especially in any bank and any ATM system. Off-Line Handwritten Signature recognition is a difficult task than On-line one because of absence of dynamic information in off-Line signature image such as angle of written style of written and so on. This paper proposed an off-Line handwritten signature recognition system with Four stages. First stage is scanning signature image and removing noise using median filter. Second stage, extract feature from each signature image using DWT technique with the advantage of multi-scale and with respect the translation, rotation and scale variations of signature image. Computing wavelet energy values from DWT details sub-bands coefficient to all person signature images using the suitable wavelet function to our database. Daubechies 20 (db20) is recognize as a suitable wavelet function with three levels analysis after a comparative study with other wavelet function. Third stage, taking the wavelet energy values as input to WNN with Morlet function as activation function in hidden layer. Finally, testing trained WNN with seen/unseen signature to evaluate our proposed system

recognition rate. A comparative study between three off-line handwritten signature systems is made (signature image pixels intensity as input to ANN, signature wavelet energy values as input to ANN and signature wavelet energy values as input to WNN). The conclusion will found that our proposed system (wavelet energy values as input to WNN) has high recognition rate.

To improve our system recognition rate, each person signature should have its own wavelet function in feature extraction stage. Genetic algorithm will be used as a searching strategy in the future work to found the suitable wavelet function to each person signature.

ACKNOWLEDGEMENTS:

The authors would like to thank Prof. Albert Swart for making his signature database available to us.

The first author would like to thank Sarah Elmetwally and Eslam Foud for their encouragements.

REFERENCES:

- [1] S. Impedovo, G. Pirolo, "Verification of Handwritten Signatures: an Overview", 14th International Conference on Image Analysis and Processing, **2007**, pp. 191-196.
- [2] M. Ammar, Y. Yoshida, T. Fulumura, "A New Effective Approach for Off-line Verification of Signatures by Using Pressure Features", Proceedings of the 8th International Conference on Pattern Recognition, **1986**, p.p 566-569.
- [3] Y. Qi, B.R. Hunt, "Signature Verification Using Global and Grid Features", Pattern Recognition, vol.27, Issue.12, **1994**, p.p 1621-1629.
- [4] V.E. Ramesh, M.N. Murty, "Off-line Signature Verification Using Genetically Optimized Weighted Features", Pattern Recognition, vol.32, Issue.2, **1999**, p.p 217-233.
- [5] P.S. Deng, H.-Y. M. Liao, C.W. Ho, and H.-R. Tyan, "Wavelet-Based Off-line Handwritten Signature Verification", Computer Vision and Image Understanding, vol.76, no.3, **1999**, p.p 173 - 190.
- [6] <http://dspace.mak.ac.ug/bitstream/123456789/599/3/karanja-evanson-mwangi-cit-masters-report.pdf>, 17-8-2010
- [7] M. Kalera, S. Srihari, and A. Xu, "Offline signature verification and identification using distance statistics", International Journal of Pattern Recognition and Artificial Intelligence. vol. 18, no.7, **2004**, pp. 1339-1360.
- [8] V. Nalwa, "Automatic on-line signature verification", Lecture Notes In Computer Science, Proceedings of the Third Asian Conference on Computer Vision, **1998**, p.p 10 - 15.
- [9] http://research.microsoft.com/pubs/69437/handwritingregistration_cvpr07.pdf, 17-8-2010.
- [10] Sing-Tze Bow, "Pattern recognition and image preprocessing", Marcel Dekker, Inc, chapter 15, **2002**.
- [11] G.Y. Chen, T.D. Bui, A. Krzyzak, "Contour-based handwritten numeral recognition using multi-wavelets and neural networks", Pattern Recognition, Vol.36, **2003**, p.p 1597 - 1604.
- [12] K R Radhika, M K Venkatesha and G N Sekhar, "Pattern Recognition Techniques in Off-line hand written signature verification - A Survey", PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY, vol. 36, ISSN 2070-3740, **2008**.
- [13] Engin Avci, Abdulkadir Sengur, Davut Hanbay, "An optimum feature extraction method for texture classification", Expert Systems with Applications: An International Journal, Published by Elsevier Ltd, Volume 36, Issue 3, **2009**, p.p 6036-6043.
- [14] <http://www.dtic.upf.edu/~xserra/cursos/TDP/referencies/Park-DWT.pdf>, 17-8-2010.
- [15] S. Sitharama Lyengar, E.C. Cho, Vir V. Phoh, "Foundations of Wavelet Networks and Application", Chapman & Hall/CRC Press LLC, chapter 4, **2002**.
- [16] Xian-Bin Wen, Hua Zhang, and Fa-Yu Wang, "A Wavelet Neural Network for SAR Image Segmentation", Sensors, Vol.9, No.9, **2009**, p.p 7509-7515.
- [17] Zhang Q. and Benveniste A, "Wavelet networks", IEEE Trans. On Neural Networks, Vol.3, **1992**, p.p 889-898.
- [18] ri Gross and Longin Jan Latecki, "Digital geometric methods in document image analysis", Pattern Recognition, Vol. 32, No.3, **1999**, pp. 407.
- [19] <http://homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm>, 24-6-2010.
- [20] Avci, E., Turkoglu, I., & Poyraz, M., "Intelligent target recognition based on wavelet packet neural network", Experts Systems with Applications: An International Journal, vol 29, Issue 1, **2005**, p.p 175-182.
- [21] Samaneh Ghandali, Mohsen Ebrahimi Moghaddam, "Off-Line Persian Signature Identification and Verification Based on Image Registration and Fusion", JOURNAL OF MULTIMEDIA, VOL. 4, NO. 3, **2009**.
- [22] A. Wahi, E. Thirumurugan "Recognition of Objects by Supervised Neural Network using Wavelet Features", First International Conference on Emerging Trends in Engineering and Technology, **2008**.
- [23] Chi-Man Pun and Moon-Chuen Lee, "Log-Polar Wavelet Energy Signatures for Rotation and Scale Invariant Texture Classification", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 25, NO. 5, **2003**.
- [24] <http://www.mathworks.com/access/helpdesk/help/toolbox/wavelet/wenergy2.html>, 17-8-2010.
- [25] Alan McCabe, Jarrod Trevathan and Wayne Read, "Neural Network-based Handwritten Signature Verification", JOURNAL OF COMPUTERS, VOL. 3, NO. 8, **2008**.
- [26] Insung Jung, and Gi-Nam Wang, "Pattern Classification of Back-Propagation Algorithm Using Exclusive Connecting Network", World Academy of Science, Engineering and Technology, VOL. 36, **2007**.
- [27] Ming Meng, Wei Sun, "Short-term Load Forecasting Based on Rough Set and Wavelet Neural Network", International Conference on Computational Intelligence and Security, **2008**.
- [28] Mohd Fazril, Zaki Ahmad, Hj. Kamaruzaman, "The Performance of Two Mothers Wavelets in Function Approximation", Journal of Mathematical Research, Vol.1, No.2, **2009**.

A Black-Box Test Case Generation Method

Nicha Kosindrdecha

Autonomous System Research Laboratory
Faculty of Science and Technology, Assumption University
Bangkok, Thailand

Jirapun Daengdej

Autonomous System Research Laboratory
Faculty of Science and Technology, Assumption University
Bangkok, Thailand

Abstract—Test case generation techniques have been researched over a long period of time. Unfortunately, while many researchers have found methods of minimizing test cases, there are still a number of important related issues that need to be researched. The primarily outstanding research issue is a large single test suite containing a huge number of test cases. Our study shows that this can lead to other two problems: unable to identify suitable test cases for execution and those test cases are lack of ability to cover domain specific requirement. Therefore, we proposed an additional requirement prioritization process during a test case generation process and an automated method to generate multiple test suites while minimizing a number of test cases from UML Use Case diagram 2.0. Our evaluation result shows that the proposed method is the most recommendation method to minimize size of test cases while maximizing ability to cover critical domain specific requirements.

Keywords—component; Test generation, testing and quality, test case generation, test generation technique and generate tests

I. INTRODUCTION

Software testing is known as a key critical phase in the software development life cycle, which account for a large part of the development effort. A way of reducing testing effort, while ensuring its effectiveness, is to generate a minimize number of test cases automatically from artifacts used in the early phases of software development. Many test case generation techniques have been proposed [2], [4], [10], [11], [12], [15], [21], [22], [42], [47], [50], mainly random, path-oriented, goal-oriented and model-based approaches. Random techniques determine a set of test cases based on assumptions concerning fault distribution. Path-oriented techniques generally use control flow graph to identify paths to be covered and generate the appropriate test cases for those paths. Goal-oriented techniques identify test cases covering a selected goal such as a statement or branch, irrespective of the path taken. There are many researchers and practitioners who have been working in generating a set of test cases based on the specifications. Modeling languages are used to get the specification and generate test cases. Since Unified Modeling Language (UML) 2.0 is the most widely used language, many researchers are using UML diagrams such as UML Use Case diagram, UML Activity diagram and UML Statechart diagram to generate test cases and this has led to model-based test case generation techniques. The study shows that model-based test generation methods (or also known as black-box test

generation) are widely-used for generating test cases in the commercial industry.

Moreover, the study [2], [4], [10], [11], [12], [15], [21], [22] shows that the primary research issue is that existing black-box test case generation methods generate a huge single test suite with a number of possible tests. The number of possible black-box tests for any non-trivial software application is extremely large. Consequently, it is unable to identify suitable test cases for execution.

Also, the study shows that the secondary research issue is that the existing black-box test case generation methods ignore critical domain specific requirements [5] during a test case generation process. These requirements are one of the most important requirements that should be addressed during test activities.

Therefore, we propose a new black-box test case generation, with requirement prioritization approach, from requirements captured as use cases, 2.0, [23], [24], [33]. A use case is the specification of interconnected sequences of actions that a system can perform, interacting with actors of the system. Use cases have become one of the favorite approaches for requirements capture. Our automated black-box approach aims to generate a minimize number of suitable test cases while reserving critical domain specific requirements. Additionally, we introduce an automated test generation method derived from UML Use Case diagram, 2.0. Our approach is developed to automatically generate many test suites based on notions announced in the latest version of UML.

The rest of the paper is organized as follow. Section 2 discusses an overview of test case generation techniques. Section 3 describes motivated research issues. Section 4 introduces a new test generation process with requirement prioritization step. Also, section 4 proposes a new black-box test generation method. Section 5 describes an experiment, measurement metrics and results. Section 6 provides the conclusion and research directions in the test case generation field. The last section represents all source references used in this paper.

II. LITERATURE REVIEW

The literature review is structured into two sections. The first section gives an overview of previous studies. The second section provides the related works

A. *An Overview of Recent Researches*

Model-based techniques are popular and most researchers have proposed several techniques. One of the reasons why those model-based techniques are popular is that wrong interpretations of complex software from non-formal specification can result in incorrect implementations leading to testing them for conformance to its specification standard [43]. A major advantage of model-based V&V is that it can be easily automated, saving time and resources. Other advantages are shifting the testing activities to an earlier part of the software development process and generating test cases that are independent of any particular implementation of the design [7].

The model-based techniques are method to generate test cases from model diagrams like UML Use Case diagram [23], [24], [33], UML Sequence diagram [7] and UML State diagram [5], [43], [22], [2], [21], [15], [32], [4]. There are many researchers who investigated in generating test cases from those diagrams. The following paragraphs show examples of model-based test generation techniques that have been proposed for a long time.

Heumann [23] presented how using use cases, derived from UML Use Case diagram 1.0, to generate test cases can help launch the testing process early in the development lifecycle and also help with testing methodology. In a software development project, use cases define system software requirements. Use case development begins early on, so real use cases for key product functionality are available in early iterations. According to the Rational Unified Process (RUP), a use case is used to describe fully a sequence of actions performed by a system to provide an observable result of value to a person or another system using the product under development." Use cases tell the customer what to expect, the developer what to code, the technical writer what to document, and the tester what to test. He proposed three-step process to generate test cases from a fully detailed use case: (a) for each use case, generate a full set of use-case scenarios (b) for each scenario, identify at least one test case and the conditions that will make it execute and (c) for each test case, identify the data values with which to test.

Ryser [24] raised the practical problems in software testing as follows: (1) Lack in planning/time and cost pressure, (2) Lacking test documentation, (3) Lacking tool support, (4) Formal language/specific testing languages required, (5) Lacking measures, measurements and data to quantify testing and evaluate test quality and (6) Insufficient test quality. They proposed their approach to resolve the above problems. Their approach is to derive test case from scenario / UML Use Case diagram 1.0 and state diagram 1.0. In his work, the generation of test cases is done in three processes: (a) preliminary test case definition and test preparation during scenario creation (b) test case generation from Statechart and from dependency charts and (c) test set refinement by application dependent strategies.

B. *Related Works*

This section provides the related works used in this paper, prioritize requirement methods. Donald Firesmith [16] addressed the purpose of requirement prioritization as follows: (a) Determine the relative necessity of the requirements.

Whereas all requirements are mandatory, some are more critical than others. For example, failure to implement certain requirements may have grave business ramifications that would make the system a failure, while others although contractually binding would have far less serious business consequences if they were not implemented or not implemented correctly (b) Help programs through negotiation and consensus building to eliminate unnecessary potential "requirements" (i.e., goals, desires, and "nice-to-haves" that do not merit the mandatory nature of true requirements) and (c) schedule the implementation of requirements (i.e., help determine what capabilities are implemented in what increment). Additionally, these researches in 1980-2008 [8], [27], [28], [29], [30], [38] reveal that there are many requirement prioritization methods such as Binary Search Tree (BST), 100-point method and Analytic Hierarchy Process (AHP)

III. RESEARCH PROBLEM

This section discusses the details of research issues related to test case generation techniques and research problems, which are motivated this study. Every test case generation technique has weak and strong points, as addressed in the literature survey. In general, referring to the literature review, the following lists major outstanding research challenges.

The first research problem is that existing test case generation methods are lack of ability to identify domain specific requirements. The study [5] shows that domain specific requirements are some of the most critical requirements required to be captured for implementation and testing, such as constraints requirements and database specific requirements. Existing approaches ignore an ability to address domain specific requirements. Consequently, software testing engineers may ignore the critical functionality related to the critical domain specific requirements. Thus, this paper introduces an approach to priority those specific requirements and generates an effective test case.

The second problem is that existing black-box test case generation techniques aim to generate a large single test suite with all possible test cases which maximize cover for each scenario. Basically, they generate a huge number of test cases which are impossible to execute given limited time and resources. As a result, those unexecuted test cases are useless and it is unable to identify suitable test cases for execution.

IV. PROPOSED METHOD

A. *Test Case Generation Process*

This section presents a new high-level process to generate a set of test cases introduced by using the above comprehensive literature review and previous works [43].

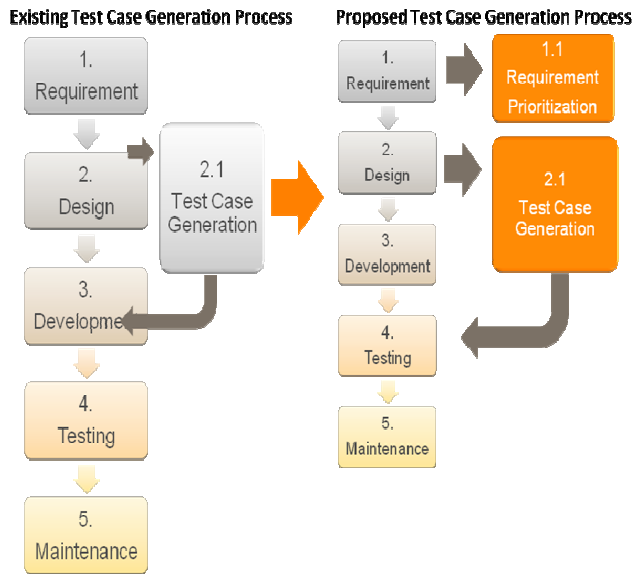


Figure 1. A Proposed Process to Generate Test Cases

From the above figure, there are two test case generation processes: existing and proposed process. The left-hand side shows an existing process to generate test cases directly from diagrams. Meanwhile, the right-hand side proposes to add an additional requirement prioritization process before generating test cases. The requirement prioritization process aims to be able to effectively handle with a large number of requirements. The objective of this process is to prioritize and organize requirements in an appropriate way in order to effectively design and prepare test cases [16], [25], [37]. There are two sub-processes: (a) classify requirements and (b) prioritize requirements.

Our study [51], [52], [53], [54] shows that a marketing perspective concentrates on two factors: customer's needs and customer satisfaction. We apply that perspective to the requirement prioritization and propose the following:

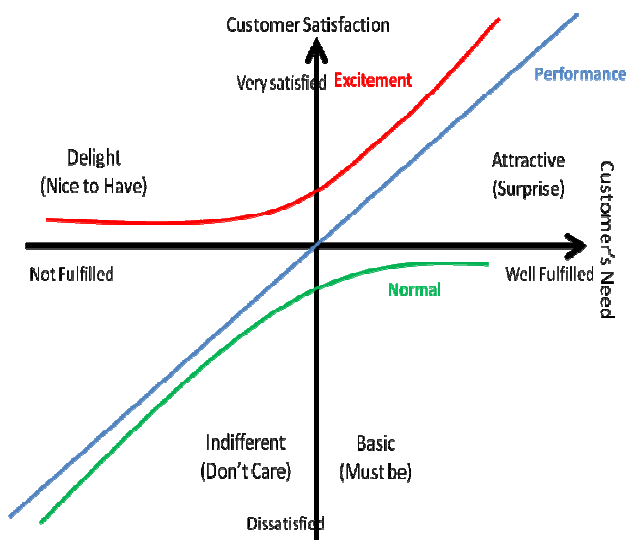


Figure 2. Classify Requirement on Marketing's Perspective

From the above figure, the horizontal axis presents a customer's need while the vertical axis represents a customer satisfaction. There are four groups of requirements based on those two factors: delight, attractive, indifferent and basic. First, the delight requirement is known as 'nice-to-have' requirement. If this requirement is well fulfilled, it will increase the customer satisfaction. Otherwise, it will not decrease the satisfaction. Second, the attractive requirement is called as 'surprise' or 'know your customer' requirement. This requirement can directly increase the customer satisfaction if it is fulfilled. Marketers and sales [53] believe that if we can deliver this kind of requirement, it will impress customers and significantly improve the customer satisfaction. Third, the indifferent requirement is a requirement that customer does not concentrate and it will not impress customers at all. In the competitive industry, this requirement may be fulfilled, but there are no any impacts to the customer satisfaction. Last, the basic requirement is a mandatory requirement that customers basically expect. Therefore, if this requirement is well delivered, it will not increase the customer satisfaction.

Furthermore, our study reveals that the requirement can be simply divided into two types: functional and non-functional requirement. Our study also presents that functional requirements can be categorized into two groups: domain specific requirement [5] (or known as constraints requirement) and behavior requirement. The following shows the requirement classification used in this paper:

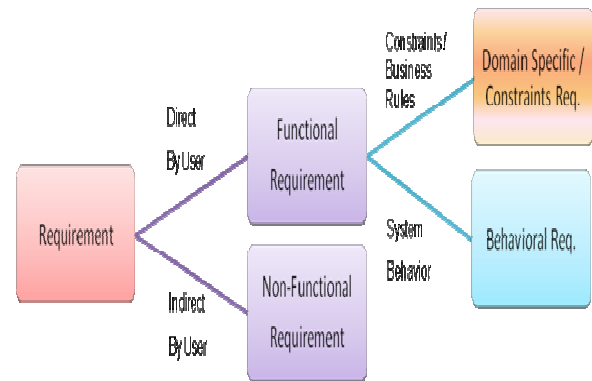


Figure 3. Classify Requirement on Software Engineer

From the above figure, functional requirement is a requirement that customers directly are able to provide. The non-functional requirement is a requirement that is given indirectly. The domain specific or constraints requirement is a requirement relative to any constraints and business rules in the software development. Meanwhile, the behavior requirement is a requirement that describes a behavior of system. Once the requirement is classified based on previous two perspectives, the next process is to prioritize requirements based on return on investment (or ROI) [51], [52], [53]. From business perspective, ROI is the most important factor to assess the important of each requirement. The following presents a ranking tree by combining those two perspectives.

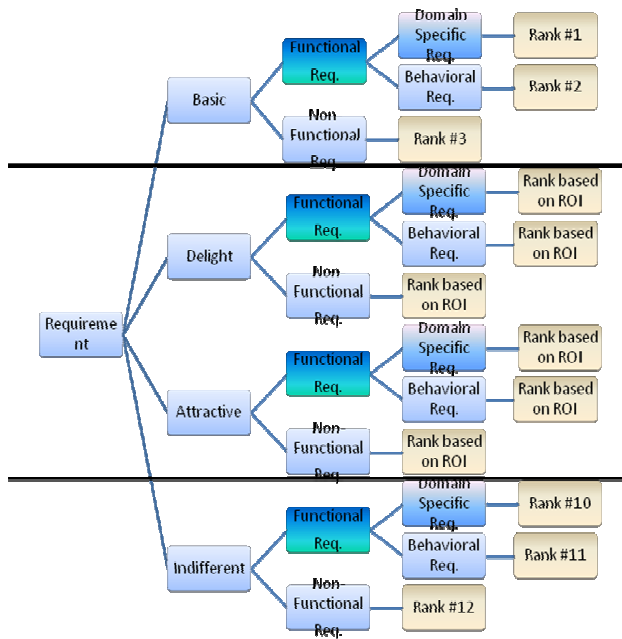


Figure 4. Requirement Prioritization Tree

From the above figure, we give the highest priority for all 'basic requirements due to the fact that they must be implemented even they do not increase the customer satisfaction. We rank the lowest priority for all 'indifferent' requirements, because customers do not concentrate on. Additionally, we prioritize both of all 'delight' and 'attractive' requirement based on ROI. In this paper, we propose to use a cost-value approach to weight and prioritize requirements. This paper proposes to use the following formula:

$$P(Req) = (Cost * CP) \quad (1)$$

Where:

- P is a prioritization value.
- Req is a requirement required to be prioritized.
- $Cost$ is a total estimated cost of coding and testing for each requirement.
- CP is an user-defined customer priority value. This value is in the range between 1 and 10. 10 is the highest priority and 1 is the lowest priority. This value aims to allow customers to identify how important of each requirement is from their perspective.

To compute the above cost for coding and testing, this paper proposes to apply the following formula:

$$Cost = (EffCode * CostCode) + (EffTest * CostTest) \quad (2)$$

Where

- $Cost$ is a total estimated cost.
- $EffCode$ is an estimated effort of coding for each requirement. The unit is man-hours.

- $CostCode$ is a cost of coding that is charged to customers. This paper applies the cost-value approach to identify the cost of coding for each requirement group (e.g. "Must-Have", "Should-Have", "Could-Have" and "Wish"). The unit is US dollar.

- $EffTest$ is an estimated effort of testing for each requirement. The unit is man-hours.

- $CostTest$ is a cost of testing that is charged to customers. The approach to identify this value is similar to $CostCode$'s approach. The unit is US dollar.

In this paper, we assumed the following in order to calculate $CostCode$ and $CostTest$. Also, this paper assumes that a standard cost for both activities is \$100 per man-hours.

- A value is 1.5 of ("Must-Have", "Should-Have") – this means that "Must-Have" requirements have one and half times cost value than "Should-Have" requirements.
- A value is 3 of ("Must-Have", "Could-Have") – this means that "Must-Have" requirements have three times cost value than "Could-Have" requirements.
- A value is 2 of ("Should-Have", "Could-Have") – this means that "Should-Have" requirements have two times cost value than "Could-Have" requirements.
- A value is approximately 3 of ("Could-Have", "Wish") – this means that "Could-Have" requirements have three times cost value than "Wish" requirements.

Therefore, the procedure of requirement prioritization process can be shortly described below:

1. Provide estimated efforts of coding and testing for each requirement.
2. Assign cost value for each requirement group based on the previous requirement classification (e.g. "Must-Have", "Should-Have", "Could-Have" and "Wish").
3. Calculate a total estimated cost for coding and testing, by using the formula (2).
4. Define a customer priority for each requirement.
5. Compute a priority value for each requirement by using the formula (1).
6. Prioritize requirements based on the higher priority value.

Once the requirements are prioritized, the next proposed step is to generate test scenario and prepare test case.

B. Test Case Generation Technique

This section presents an automated test scenario generation derived from UML Use Case diagram 2.0. The big different between UML Use Case diagram 1.0 and 2.0 is a package notion that can group each use case into each package.

The following shows an example of UML Use Case diagram 2.0.

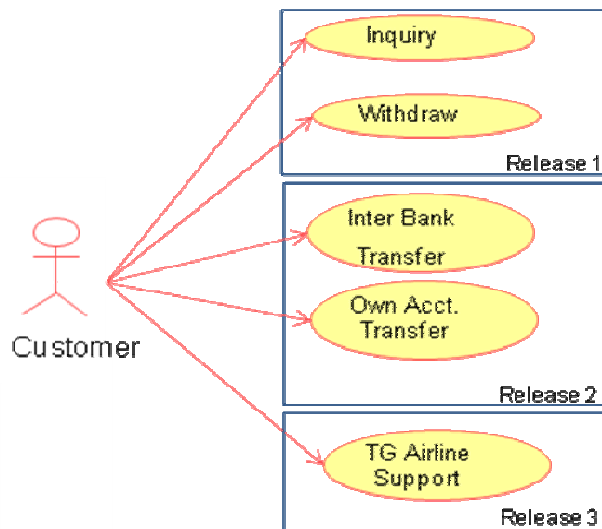


Figure 5. An Example of UML Use Case Diagram 2.0

From the above figure, the new notion in UML Use Case diagram 2.0 is a package that is used for grouping each function. There are three packages or releases. Each release contains different functional requirement. The first release contains two functions: inquiry and withdraw. The second release is composed of: transfer own account and transfer to other banks. The last release has only one function to support Thai (TG) airline tickets.

Our approach aims to generate three test suites to cover the above three packages while existing test case generation techniques do not concentrate on. The first test suite is developed for: inquiry and withdraw functions. The second test suite is used for transferring own banks and other banks. The last suite aims to a TG airline ticket support.

The approach is built based on Heumann's algorithm [23]. The limitation of our approach is to ensure that all use cases are fully dressed. The fully dressed use case is a use case with the comprehensive of information, as follows: use case name, use case number, purpose, summary, pre-condition, post-condition, actors, stakeholders, basic events, alternative events, business rules, notes, version, author and date.

The proposed method contains four steps, as follows: (a) extract use case diagram (b) generate test scenario (c) prepare test data and (d) prepare other test elements. These steps can be shortly described as follows:

1. The first step is to extract the following information from fully dressed use cases: (a) use case number (b) purpose (c) summary (d) pre-condition (e) post-condition (f) basic event and (g) alternative events. This information is called use case scenario in this paper. The example fully dressed use cases of ATM withdraw functionality can be found as follows:

TABLE I. EXAMPLE FULLY DRESSED USE CASE

Use	Use	Summar	Basic	Alternati	Busine
-----	-----	--------	-------	-----------	--------

Case Id	Case Name	y	Event	ve Events	ss Rules
UC-001	Withdraw	To allow bank's customers to withdraw money from ATM machines anywhere in Thailand.	1. Insert Card 2. Input PIN 3. Select Withdraw 4. Select A/C Type 5. Input Balance 6. Get Money 7. Get Card	1. Select Inquiry 2. Select A/C Type 3. Check Balance	(a) Input amount <= Outstanding Balance (b) Fee charge if using different ATM machines
UC-002	Transfer	To allow users to transfer money to other banks in Thailand from all ATM machines	1. Insert Card 2. Input PIN 3. Select Transfer 4. Select bank 5. Select "To" account 6. Select A/C Type 7. Input Amount 8. Get Receipt 9. Get Card	1. Select Inquiry 2. Select A/C Type 3. Check Balance	Amount <= 50,000 baht

The above use cases can be extracted into the following use case scenarios:

TABLE II. EXTRACTED USE CASE SCENARIOS

Scenario Id	Summary	Basic Scenario
Scenario-001	To allow bank's customers to withdraw money from ATM machines anywhere in Thailand.	1. Insert Card 2. Input PIN 3. Select Withdraw 4. Select A/C Type 5. Input Balance 6. Get Money 7. Get Card

Scenario-002	To allow bank's customers to withdraw money from ATM machines anywhere in Thailand.	1. Insert Card 2. Input PIN 3. Select Inquiry 4. Select A/C Type 5. Check Balance 6. Select Withdraw 7. Select A/C Type 8. Input Balance 9. Get Money 10. Get Card
Scenario-003	To allow users to transfer money to other banks in Thailand from all ATM machines	1. Insert Card 2. Input PIN 3. Select Transfer 4. Select bank 5. Select "To" account 6. Select A/C Type 7. Input Amount 8. Get Receipt 9. Get Card
Scenario-004	To allow users to transfer money to other banks in Thailand from all ATM machines	1. Insert Card 2. Input PIN 3. Select Inquiry 4. Select A/C Type 5. Check Balance 6. Select Transfer 7. Select bank 8. Select "To" account 9. Select A/C Type 10. Input Amount 11. Get Receipt 12. Get Card

TS-002	To allow bank's customers to withdraw money from ATM machines anywhere in Thailand.	1. Insert Card 2. Input PIN 3. Select Inquiry 4. Select A/C Type 5. Check Balance 6. Select Withdraw 7. Select A/C Type 8. Input Balance 9. Get Money 10. Get Card
TS-003	To allow users to transfer money to other banks in Thailand from all ATM machines	1. Insert Card 2. Input PIN 3. Select Transfer 4. Select bank 5. Select "To" account 6. Select A/C Type 7. Input Amount 8. Get Receipt 9. Get Card
TS-004	To allow users to transfer money to other banks in Thailand from all ATM machines	1. Insert Card 2. Input PIN 3. Select Inquiry 4. Select A/C Type 5. Check Balance 6. Select Transfer 7. Select bank 8. Select "To" account 9. Select A/C Type 10. Input Amount 11. Get Receipt 12. Get Card

- The second step is to automatically generate test scenarios from the previous use case scenarios [23]. From the above table, we automatically generate the following test scenarios:

TABLE III. GENERATED TEST SCENARIOS

Test Scenario Id	Summary	Basic Scenario
TS-001	To allow bank's customers to withdraw money from ATM machines anywhere in Thailand.	1. Insert Card 2. Input PIN 3. Select Withdraw 4. Select A/C Type 5. Input Balance 6. Get Money 7. Get Card

- The next step is to prepare test data. This step allows to manually prepare an input data for each scenario.
- The last step is to prepare other test elements, such as expected output, actual output and pass / fail status

V. EVALUATION

The section describes the experiments design, measurement metrics and results.

A. Experiments Design

- Prepare Experiment Data. Before evaluating the proposed methods and other methods, preparing experiment data is required. In this step, 50 requirements and 50 use case scenarios are randomly generated.
- Generate Test Scenario and Test Case. A comparative evaluation method has been made among the proposed test scenario algorithm, Heumann's technique Jim [23], Ryser's method [24], Nilawar's algorithm [33] and the proposed method presented in the previous section. It

is included a prioritization requirement algorithm prior to generate a set of test scenarios and test cases.

3. Evaluate Results. In this step, the comparative generation methods are executed by using 50 requirements and 50 use case scenarios. These methods are also executed for 10 times in order to find out the average percentage of critical domain requirement coverage, a size of test cases and total generation time. In total, there are 500 requirements and 500 use case scenarios executed in this experiment.

The following tables present how to randomly generate data for requirements and use case scenarios respectively.

TABLE IV. GENERATE RANDOM REQUIREMENTS

Attribute	Approach
Requirement ID	Randomly generated from the following combination: Req + <i>Sequence Number</i> . For example, Req1, Req2, Req3, ..., ReqN.
Description	Randomly generated from the following combination: Des + <i>Sequence Number same as Requirement ID</i> . For example, Des1, Des2, Des3, ..., DesN.
Type of Requirement	Randomly selected from the following values: Functional AND Non-Functional.
MoSCoW Criteria	Randomly selected from the following values: Must Have (M), Should Have (S), Could Have (C) and Won't Have (W)
Is it a critical requirement (Y/N)?	Randomly selected from the following values: True (Y) and False (N)

TABLE V. GENERATE RANDOM USE CASE SCENARIO

Attribute	Approach
Use case ID	Randomly generated from the following combination: uCase + <i>Sequence Number</i> . For example, <i>uCase₁</i> , <i>uCase₂</i> , ..., <i>uCase_n</i> .
Purpose	Randomly generated from the following combination: Pur + <i>Sequence Number same as Use case ID</i> . For example, <i>Pur₁</i> , <i>Pur₂</i> , ..., <i>Pur_n</i> .
Pre-condition	Randomly generated from the following combination: pCon + <i>Sequence Number same as Use case ID</i> . For example, <i>pCon₁</i> , <i>pCon₂</i> , ..., <i>pCon_n</i> .
Basic Scenario	Randomly generated from the following combination: uCase + <i>Sequence Number</i> . For example,

	<i>basic₁</i> , <i>basic₂</i> , ..., <i>basic_n</i> .
--	---

B. Measurement Metrics

The section lists the measurement metrics used in the experiment. This paper proposes to use three metrics, which are: (a) size of test cases (b) total time and (c) percentage of critical domain requirement coverage. The following describe the measurement in details.

1. A Number of Test Cases: This is the total number of generated test cases, expressed as a percentage, as follows:

$$\% \text{ Size} = (\# \text{ Size} / \# \text{ of Total Size}) * 100 \quad (3)$$

Where:

- *% Size* is a percentage of the number of test cases.
 - *# of Size* is a number of test cases.
 - *# of Total Size* is the maximum number of test cases in the experiment, which is assigned 1,000.
2. A Domain Specific Requirement Coverage: This is an indicator to identify the number of requirements covered in the system, particularly critical requirements, and critical domain requirements [5]. Due to the fact that one of the goals of software testing is to verify and validate requirements covered by the system, this metric is a must. Therefore, a high percentage of critical requirement coverage is desirable.

It can be calculated using the following formula:

$$\% \text{ CRC} = (\# \text{ of Critical} / \# \text{ of Total}) * 100 \quad (4)$$

Where:

- *% CRC* is the percentage of critical requirement coverage.
 - *# of Critical* is the number of critical requirements covered.
 - *# of Total* is the total number of requirements.
3. Total Time: This is the total number of times the generation methods are run in the experiment. This metric is related to the time used during the testing development phase (e.g. design test scenario and produce test case). Therefore, less time is desirable.

It can be calculated using the following formula:

$$\text{Total} = P\text{Time} + C\text{Time} + R\text{Time} \quad (5)$$

Where:

- *Total* is the total amount of times consumed by running generation methods.
- *PTime* is the total amount of time consumed by preparation before generating test cases.
- *CTime* is the time to compile source code / binary code in order to execute the program.
- *RTime* is the total time to run the program under this experiment.

C. Results and Discussion

This section discusses an evaluation result of the above experiment. This section presents a graph that compares the above proposed method to other three existing test case generation techniques, based on the following measurements: (a) size of test cases (b) critical domain coverage and (c) total time. Those three techniques are: (a) Heumann's method (b) Ryser's work and (c) Nilawar's approach. There are two dimensions in the following graph: (a) horizontal and (b) vertical axis. The horizontal represents three measurements whereas the vertical axis represents the percentage value.

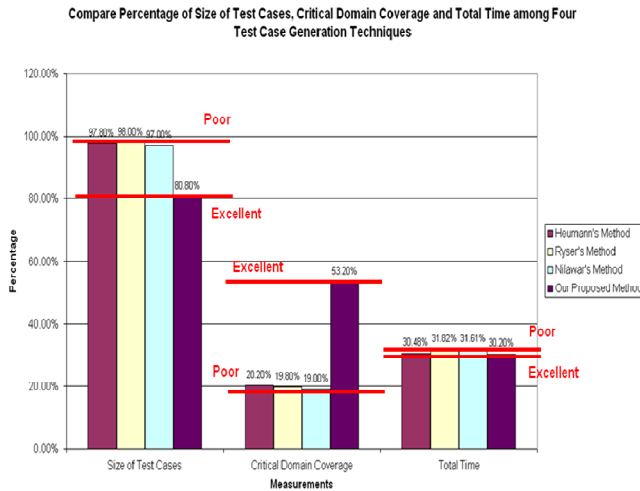


Figure 6. An Evaluation Result of Test Generation Methods

The above graph shows that the above proposed method generates the smallest set of test cases. It is calculated as 80.80% whereas the other techniques are computed over 97%. Those techniques generated a bigger set of test cases, than a set generated by the proposed method. The literature review reveals that the smaller set of test cases is desirable. Also, the graph shows that the proposed method consumes the least total time during a generation process, comparing to other techniques. It used only 30.20%, which is slightly less than others. Finally, the graph presents that the proposed method is the best technique to coverage critical domains. Its percentage is much greater than other techniques' percentage, over 30%.

From the above figure, this study determines and ranks the above comparative methods into five ranking: 5-Excellent, 4-Very good, 3-Good, 2-Normal and 1-Poor. This study uses a maximum and minimum value to find an interval value for ranking those methods.

For a number of test cases, the maximum and minimum percentage is 98% and 80.80%. The different between maximum and minimum value is 17.2%. An interval value is equal to a result of dividing the different values by 5. As a result, the interval value is approximately 3.4. Thus, it can be determined as follows: 5-Excellent (since 80.80% to 84.2%), 4-Very good (between 84.2% and 87.6%), 3-Good (between 87.6% and 91%), 2-Normal (between 91% and 94.4%) and 1-Poor (from 94.4% to 97.8%).

For an ability to cover critical domain specific requirement, the maximum and minimum percentage is 53.20% and 19%. The different value is 34.2%. The interval value is 6.84. Therefore, it can be determined as follows: 5-Excellent (since 46.36% to 53.2%), 4-Very good (between 39.52% and 46.36%), 3-Good (between 32.68% and 39.52%), 2-Normal (between 25.84% and 32.68%) and 1-Poor (from 19% to 25.84%).

For a total time, the maximum and minimum percentage is 31.82% and 30.20%. The different between maximum and minimum value is 1.62%. An interval value is equal to a result of dividing the different values by 5. As a result, the interval value is 0.324. Thus, it can be determined as follows: 5-Excellent (since 30.2% to 30.524%), 4-Very good (between 30.524% and 30.848%), 3-Good (between 30.848% and 31.172%), 2-Normal (between 31.172% and 31.496%) and 1-Poor (from 31.496% to 31.82%).

Therefore, the experiment result of those comparative methods can be shown below:

TABLE VI. A COMPARISON OF TEST CASE REDUCTION METHODS

Algorithm	A Number of Test Cases	Cover Critical Domain Specific Req.	Total Time
Heumann's Method	1	1	5
Ryser's Method	1	1	1
Nilawar's Method	1	1	1
Our Proposed Method	5	5	5

In the conclusion, the proposed method is the best to generate the smallest size of test cases with the maximum of critical domain coverage and the least time consumed in the generation process.

VI. CONCLUSION

In this paper, we introduced a new test case generation method and process, with an additional requirement prioritization process. The approach inserts an additional process to ensure that all domain specific requirements are captured during the test case generation. Also, the approach is developed to minimize a number of test cases in order to be able to select suitable test cases for execution. Additionally, we proposed an automated approach to generate test cases from fully described UML use cases, version 2.0. Our generated method can generate many test suites derived from UML Use Case diagram, 2.0. Existing test case generation methods generate only a large single test suite that contains a lot of numbers of test cases.

Furthermore, we conducted an evaluation experiment with a random requirements and fully described use cases. Our

evaluation result reveals that the proposed method is the most recommendation automated test case generation methods for maximizing critical domain requirement coverage. Also, the result present that the proposed method is one of best methods to minimize a number of test cases.

The future research, we plan to enhance an ability to prioritize requirements and conduct a large experiment for a large system development

REFERENCES

- [1] Ahl, V. "An Experimental Comparison of Five Prioritization Methods" Master's Thesis, School of Engineering, Blekinge Institute of Technology, Ronneby, Sweden, 2005.
- [2] Alessandra Cavarra, Charles Crichton, Jim Davies, Alan Hartman, Thierry Jeron and Laurent Mounier. "Using UML for Automatic Test Generation." Oxford University Computing Laboratory, Tools and Algorithms for the Construction and Analysis of Systems, TACAS'2000, 2000.
- [3] Amaral. "A.S.M.S. Test case generation of systems specified in Statecharts." M.S. thesis – Laboratory of Computing and Applied Mathematics, INPE, Brazil, 2006.
- [4] Annelises A. Andrews, Jeff Offutt and Roger T. Alexander. "Testing Web Applications. Software and Systems Modeling.", 2004.
- [5] Avik Sinha, Ph.D and Dr. Carol S. Smidts. "Domain Specific Test Case Generation Using Higher Ordered Typed Languages fro Specification." Ph. D. Dissertation, 2005.
- [6] A. Bertolino. "Software Testing Research and Practice." 10th International Workshop on Abstract State Machines (ASM'2003), Taormina, Italy, 2003.
- [7] A.Z. Javed, P.A. Strooper and G.N. Watson "Automated Generation of Test Cases Using Model-Driven Architecture." Second International Workshop on Automation of Software Test (AST'07), 2007.
- [8] Beck, K. & Andres, C. "Extreme Programming Explained: Embrace Change", 2nd ed. Boston, MA: Addison-Wesley, 2004.
- [9] Boehm, B. & Ross, R. "Theory-W Software Project Management: Principles and Examples." IEEE Transactions on Software Engineering 15, 4: 902-916, 1989.
- [10] B.M. Subraya, S.V. Subrahmanya "Object driven performance testing in Web applications." in: Proceedings of the First Asia-Pacific Conference on Quality Software (APAQS'00), pp. 17-26, Hong Kong, China, 2000.
- [11] Chien-Hung Liu, David C. Kung, Pei Hsia and Chih-Tung Hsu "Object-Based Data Flow Testing of Web Applications." Proceedings of the First Asia-Pacific Conference on Quality Software (APAQS'00), pp. 7-16, Hong Kong, China, 2000.
- [12] C.H. Liu, D.C. Kung, P. Hsia, C.T. Hsu "Structural testing of Web applications." in: Proceedings of 11th International Symposium on Software Reliability Engineering (ISSRE 2000), pp. 84-96, 2000.
- [13] Davis, A. "The Art of Requirements Triage." IEEE Computer 36, 3 p: 42-49, 2003.
- [14] Davis, A. "Just Enough Requirements Management: Where Software Development Meets Marketing." New York: Dorset House (ISBN 0-932633-64-1), 2005.
- [15] David C. Kung, Chien-Hung Liu and Pei Hsia "An Object-Oriented Web Test Model for Testing Web Applications." In Proceedings of the First Asia-Pacific Conference on Quality Software (APAQS'00), page 111, Los Alamitos, CA, 2000.
- [16] Donald Firesmith "Prioritizing Requirements. Journal of Object Technology", Vol.3, No8, 2004.
- [17] D. Harel "On visual formalisms." Communications of the ACM, vol. 31, no. 5, pp. 514-530, 1988.
- [18] D. Harel. "Statecharts: A Visual Formulation for Complex System." Sci.Comput. Program. 8(3):232-274, 1987.
- [19] Flippo Ricca and Paolo Tonella "Analysis and Testing of Web Applications." Proc. of the 23rd International Conference on Software Engineering, Toronto, Ontario, Canada. pp.25-34, 2001.
- [20] Harel, D. "Statecharts: a visual formalism for complex system." Science of Computer Programming, v. 8, p. 231-274, 1987.
- [21] Hassan Reza, Kirk Ogaard and Amarnath Malge "A Model Based Testing Technique to Test Web Applications Using Statecharts." Fifth International Conference on Information Technology, 2008.
- [22] Ibrahim K. El-Far and James A. Whittaker "Model-based Software Testing", 2000.
- [23] Jim Heumann "Generating Test Cases From Use Cases." Rational Software, 2001.
- [24] Johannes Ryser and Martin Glinz "SCENT: A Method Employing Scenarios to Systematically Derive Test Cases for System Test", 2000.
- [25] Karl E. Wiegers "First Things First: Prioritizing Requirements." Published in Software Development, 1999.
- [26] Karlsson, J. "Software Requirements Prioritizing." Proceedings of the Second International Conference on Requirements Engineering (ICRE'96). Colorado Springs, CO, April 15-18, 1996. Los Alamitos, CA: IEEE Computer Society, p 110-116, 1996.
- [27] Karlsson, J. "Towards a Strategy for Software Requirements Selection." Licentiate. Thesis 513, Linköping University, 1995.
- [28] Karlsson, J. & Ryan, K. "A Cost-Value Approach for Prioritizing Requirements." IEEE Software September/October, p67-75, 1997.
- [29] Leffingwell, D. & Widrig, D. "Managing Software Requirements: A Use Case Approach", 2nd ed. Boston, MA: Addison-Wesley, 2003.
- [30] Leslie M. Tierstein "Managing a Designer / 2000 Project." NYOUG Fall'97 Conference, 1997.
- [31] L. Brim, I. Cerna, P. Varekova, and B. Zimmerova "Component-interaction automata as a verification oriented component-based system specification." In: Proceedings (SAVCBS'05), pp. 31-38, Lisbon, Portugal, 2005.
- [32] Mahnaz Shams, Diwakar Krishnamurthy and Behrouz Far "A Model-Based Approach for Testing the Performance of Web Applications." Proceedings of the Third International Workshop on Software Quality Assurance (SOQUA'06), 2006.
- [33] Manish Nilawar and Dr. Sergiu Dascalu "A UML-Based Approach for Testing Web Applications." Master of Science with major in Computer Science, University of Nevada, Reno, 2003.

- [34] Moisiadis, F. "Prioritising Scenario Evolution." International Conference on Requirements Engineering (ICRE 2000), 2000.
- [35] Moisiadis, F. "A Requirements Prioritisation Tool." 6th Australian Workshop on Requirements Engineering (AWRE 2001). Sydney, Australia, 2001.
- [36] M. Prasanna S.N. Sivanandam R.Venkatesan R.Sundarrajan "A Survey on Automatic Test Case Generation." Academic Open Internet Journal, 2005.
- [37] Nancy R. Mead "Requirements Prioritization Introduction." Software Engineering Institute, Carnegie Mellon University, 2008.
- [38] Park, J.; Port, D.; & Boehm B. "Supporting Distributed Collaborative Prioritization for Win-Win Requirements Capture and Negotiation." Proceedings of the International Third World Multi-conference on Systemics, Cybernetics and Informatics (SCI'99) Vol. 2. 578-584, Orlando, FL, July 31-August 4, 1999. Orlando, FL: International Institute of Informatics and Systemic (IIIS), 1999.
- [39] Rajib "Software Test Metric." QCON, 2006.
- [40] Robert Nilsson, Jeff Offutt and Jonas Mellin "Test Case Generation for Mutation-based Testing of Timeliness.", 2006.
- [41] Saaty, T. L. "The Analytic Hierarchy Process." New York, NY: McGraw-Hill, 1980.
- [42] Shengbo Chen, Huaikou Miao, Zhongsheng Qian "Automatic Generating Test Cases for Testing Web Applications." International Conference on Computational Intelligence and Security Workshops, 2007.
- [43] Valdivino Santiago, Ana Silvia Martins do Amaral, N.L. Vijaykumar, Maria de Fatima, Mattiello-Francisco, Eliane Martins and Odnei Cuesta Lopes "A Practical Approach for Automated Test Case Generation using Statecharts", 2006.
- [44] Vijaykumar, N. L.; Carvalho, S. V.; Abdurahiman, V. "On proposing Statecharts to specify performance models." International Transactions in Operational Research, 9, 321-336, 2002.
- [45] Wiegers, K. "E. Software Requirements", 2nd ed. Redmond, WA: Microsoft Press, 2003.
- [46] Xiaoping Jia, Hongming Liu and Lizhang Qin "Formal Structured Specification for Web Application Testing". Proc. of the 2003 Midwest Software Engineering Conference (MSEC'03). Chicago, IL, USA. pp.88-97, 2003.
- [47] Yang, J.T., Huang, J.L., Wang, F.J. and Chu, W.C. "Constructing an object-oriented architecture for Web application testing." Journal of Information Science and Engineering 18, 59-84, 2002.
- [48] Ye Wu and Jeff Offutt "Modeling and Testing Web-based Applications", 2002.
- [49] Ye Wu, Jeff Offutt and Xiaochen "Modeling and Testing of Dynamic Aspects of Web Applications, Submitted for publication." Technical Report ISE-TR-04-01, www.ise.gmu.edu/techreps/, 2004.
- [50] Zhu, H., Hall, P., May, J. "Software Unit Test Coverage and Adequacy." ACM Comp. Survey 29(4), pp 366~427, 1997.
- [51] Kano Noriaki, Nobuhiku Seraku, Fumio Takahashi, Shinichi Tsuji. "Attractive Quality and Must-Be Quality." Journal of the Japanese Society for Quality Control. 14(2), pp 39~48, 1984.
- [52] Cadotte, Ernest R., Turgeon, Normand "Dissatisfiers and Satisfiers: Suggestions from Consumer Complaints and Compliments." Journal of Consumer Satisfaction, Dissatisfactions and Complaining Behavior. 1, pp 74~79, 1988.
- [53] Brandt, D. Randall "How service marketers can identify value-enhancing service elements." Journal of Services Marketing. 2(3), pp 35~41, 1988.
- [54] Herzberg, Frederick, Mausner, B., Snyderman, B.B. "The motivation to work." New York: Wiley, 2nd edition, 1959.

White-Box Test Reduction Using Case-Based Maintenance

Siripong Roongruangsuwan

Autonomous System Research Laboratory
Faculty of Science and Technology, Assumption University
Bangkok, Thailand

Jirapun Daengdej

Autonomous System Research Laboratory
Faculty of Science and Technology, Assumption University
Bangkok, Thailand

Abstract— Software testing has been proven that it takes around 50-70% of the costs associated with the large development of commercial software systems. Many reduction techniques have been proposed to reduce costs. Unfortunately, the cost is usually over budget and those techniques are failed to reasonably control costs. The primarily outstanding research issues, motivated this study, are a large number of redundancy test cases and a decrease of ability to detect faults. To resolve these issues, this paper proposes new deletion algorithms to minimize a number of white-box test cases, while maximizing an ability to reveal faults, by using a concept of case-based maintenance. Our evaluations have shown that the proposed techniques can significantly reduce a number of unnecessary test cases while preserving the capability of fault detection.

Keywords—component; Test reduction, test case reduction, deletion method, case based maintenance and test case deletion

I. INTRODUCTION

Software Testing is an empirical investigation conducted to provide stakeholders with information about the quality of the product or service under test [13], with respect to the context in which it is intended to operate. Software Testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks of implementation of the software. The software testing techniques include the process of executing a program or application with the intent of finding software bugs. It can also be stated as the process of validating and verifying that software meets the business and technical requirements that guided its design and development, so that it works as expected. Software Testing can be implemented at any time in the development process; however, the most test effort is employed after the requirements have been defined and coding process has been completed.

Many researchers [6], [7], [8], [9], [10], [19], [24], [25], [26], [28], [30], [36], [37], [39] have proven that these test case reduction methods can reserve the ability to reveal faults. However, there are many outstanding research issues in this area. The motivated research issues are: a large number of test cases particularly redundancy test cases and a decrease of ability to detect faults. The study shows that test case reduction methods have been researched over a long period of time, such as test case prioritization, random approach and coverage-based

test case reduction techniques. Also, the study reveals that coverage-based approaches are wildly used and researched. Therefore, we concentrate on an approach to reduce test cases based on the coverage factor. Many coverage factors have been proposed over a long period of time. Unfortunately, existing factors and test case reduction methods ignore the complexity and impact of test cases. Thus, we propose to reduce a number of test cases by considering both of test case complexity and impact.

Our study [5] shows that one of effective approaches that significantly reduce a number of redundancy test cases is to apply the concept of artificial intelligent. There are many artificial intelligent concepts, such as neural network, fuzzy logic, learning algorithms and case-based reasoning (CBR). CBR is one of the most popular and actively researched areas in the past. The researches [4], [5], [26], [35] show that CBR has identical problems as same as software testing topic, such as a huge number of redundancy cases and a decrease of system's ability to resolve problems.

Fundamentally, there are four steps in the CBR system, which are: retrieve, reuse, revise and retain. These steps can lead to a serious problem of uncontrollably growing cases in the system. However, the study shows that there are many proposed techniques in order to control a number of cases in the CBR system, such as add algorithms, deletion algorithms and maintenance approaches. CBR have been investigated by CBR researchers in order to ensure that only small amounts of efficient cases are stored in the case base.

The previous work [27] shows that deletion algorithms are the most popular and effective approaches to maintain a size of the CBR system. There are many researchers have proposed several deletion algorithms [4], [20], [35], such as random method, utility approach and footprint algorithm. These algorithms aim to: (a) remove all redundancy or unnecessary cases (b) minimize a number of cases and (c) maintain the ability of solving problems. Nevertheless, each technique has strength and weakness. Some methods are suitable for removing cases. Some methods are perfectly suitable for reducing time. Some may be used for reserving the problem solving capability. Eventually, the previous work [27] discovered several effective methods (e.g. confidential case filtering method, coverage value algorithm and confidential coverage approach) to remove those cases, minimize size of

CBR and reduce amount of time, while preserving the ability of CBR system's problem solving skill. Therefore, this paper applies those effective deletion techniques to resolve the problems of software testing.

In the light of software testing, the proposed techniques focus on how to maintain the test case while maintaining the capability of fault detection. It is appear that test cases in this paper are treated as cases in the CBR system. Also, there is an assumption that a given set of test cases are generated by a path-oriented test case generation technique. The path-oriented technique is widely used for a white-box testing that derives test cases from available source code.

Section 2 discusses an overview of test case reduction techniques and approach to maintain CBR. Section 3 provides a definition of terminologies used in this paper. Section 4 lists the outstanding research issues motivated this study. Section 5 proposes deletion algorithms using the concept of CBR. Section 6 describes an evaluation method and discusses an evaluation result. The last section represents all source references used in this paper.

II. LITERATURE REVIEW

This section describes an overview of test case reduction techniques and the concept of case based maintenance. The following describes those two areas in details.

A. Test Case Reduction Techniques

This section discusses and organizes test case reduction (or TCR) techniques researched in 1995-2006. This study shows that there are many researchers who proposed a method to reduce unnecessary test cases (also known as redundancy test cases), like Offutt [2], Rothermel [8], McMaster [25] and Samph [31]. These techniques aim to remove and minimize a size of test cases while maintaining the ability to detect faults. The literature review [6], [7], [8], [9], [10], [11], [19], [24], [25], [36], [37], [39] shows that there are two types of reduction techniques, which are: (a) pre-process and (b) post-process. First, the pre-process is a process that immediately reduces a size of test cases after generating. Typically, it is occurred before regression testing phase. Second, the post-process is a process that maintains and removes unnecessary test cases, after running the first regression testing activities. Although these techniques can reduce the size of test cases, but the ability to reveal faults seems slightly to be dropped. However, Jefferson Offutt [5] and Rothermel [6], [7], [8], [9], [10], [19], [20], [21], [32] has proven that these test case reduction techniques have many benefits, particularly during the regression testing phase, and most of reduction techniques can maintain an acceptable rate of fault detection. The advantages of these techniques are: (a) to spend less time in executing test cases, particularly during the regression testing phase (b) to significantly reduce time and cost of manually comparing test results and (c) to effectively manage the test data associated with test cases. This study proposes a new "2C" classification of test case reduction techniques, classified based on their characteristics, as follows: (a) coverage-based techniques and (b) concept analysis-based techniques.

B. Case-Based Maintenance (CBM)

Due to the CBR's life cycle [16], the case base size grows rapidly. That is caused a serious problem directly, for instance, duplicate data, inconsistency data, incorrect data, and an expense of searching for an appropriate case in a large case base size. CBR can be classified as one of the Artificial Intelligence algorithms. CBR solves new problem by retrieving the similar case from the existing case base and then adapts the existing case according to the target problem. Over the time, CBR is growing. When the uncontrollable case-based growth is occurred, the performance of CBR is decreasing. Therefore, the maintenance process is required in order to preserve or improve the performance of the system. The process of maintaining CBR is called CBM. David C. Wilson [8] presented the overall concepts of CBR and case based maintenance. This paper focused on the case based maintenance (CBM) approach in term of the framework. In other words, this paper described the type of data collection and how the case based maintenance works. There were so many policies for CBM, for example, addition, deletion, and retain.

"CBM was defined as the process of refining a CBR system's case-base to improve the system's performance. It implements policies for revising the organization or contents (representation, domain content, accounting information, or implementation) of the case-base in order to facilitate future reasoning for a particular set of performance objectives."

These studies [2], [3], [4], [16], [17], [27], [35] reveal that several deletion algorithms have been proposed. For example, a random approach (RD), utility deletion algorithm (UD), footprint deletion algorithm (FD), footprint utility deletion algorithm (FUD) and iterative case filtering algorithm (ICF).

RD is the simplest approach, which removes the case randomly. UD deletes the case that has minimum utility value. Footprint algorithm uses the competence model and removes the auxiliary case from the system. FUD is a hybrid approach between Utility algorithm and Footprint algorithm, and is concerned with the competence model and the utility value. Finally, ICF focuses on the case, which the reachability set is greater than the coverage set [16], [27].

III. DEFINITION

This section describes a definition of CBR terminologies used in the software testing area.

TABLE I. DEFINITIONS OF CBR FOR SOFTWARE TESTING

Element	CBR	Software Testing
Coverage set	Coverage Set is the set of target problems, which it can be used to solve successfully [4].	Coverage set means a set of stages / elements, which they can be used to test successfully and reveal faults.
Reachability set	Reachability Set is the set of case bases that can be used to solve	Reachability set means a set of test cases that can be used to reveal

	the target problem [4].	faults.
Competence set	Competence is the range of the target problem that can be solved successfully [4].	Competence is the range of ability to reveal faults that can be used to test successfully.
Auxiliary set	Auxiliary Case is a case that does not have a direct effect on the competence of a system when it is deleted [4].	Auxiliary case is a test case that does not have a direct effect on the competence of a system when it is removed.
Pivot set	Pivotal Case is the case that does have a direct effect on the competence of a system if it is deleted [1], [29].	Pivotal case is a test case that does have a direct effect on the ability to reveal faults if it is deleted.

IV. RESEARCH PROBLEM

This section discusses the details of research issues motivated this study. The literature review reveals that [13], [23], [24], [25], [31], [38] those research issues are: (a) a large number of redundancy test cases and (b) a decrease of the ability to reveal faults. These research issues can be elaborated in details as follows:

First, the literature review shows that redundancy test cases are test cases tested by multiple test cases. Many test cases that are designed to test the same things (e.g. same functions, same line of code or same requirements) are duplicated. Those duplicated tests are typically occurred during testing activities, particularly during regression testing activities [13], [23], [24], [25], [31], [38]. Those duplicated tests can be eventually removed in order to minimize time and cost to execute tests.

The following shows an example of redundancy test cases.

Id	Description	Input	Sequence	Expected Result	Actual Result	Status Pass / Fail
001	Withdraw Money from ATM	PIN, Amount	1. Insert ATM Card 2. Insert PIN 3. Select "Withdraw" 4. Select A/C Type 5. Identify Amount 6. Click "OK" 7. Receive Money 8. Receive Card	Money is returned Balance is calculated. ATM Card is returned.	---	---
002	Inquiry & Withdraw Money from ATM	PIN, Amount	1. Insert ATM Card 2. Insert PIN 3. Select "Inquiry" 4. Select A/C Type 5. Click "OK" 6. Select "Withdraw" 7. Select A/C Type 8. Identify Amount 9. Click "OK" 10. Receive Money 11. Receive Card	Current balance is displayed. Money is returned Balance is calculated. ATM Card is returned.	---	---
...						

Figure 1. An Example of Control Flow Graph

From the above figure, there are two test cases, with the duplicated sequence and expected result, designed to test a "withdraw" feature in ATM machine. The sequence of the first test case is: (a) insert ATM card (b) insert PIN (c) select "withdraw" (d) select account type (e) identify an amount (f) click "OK" button (g) receive money and (h) receive card. The sequence of the second test case is similar to the first one. However, the additional step in the second case is to inquiry a balance amount before withdrawing the money. Therefore, it is appear that the first test case is a part of the second test case. We call the first test case as a redundancy test case.

The study shows that there are many proposed methods to delete those duplicated test cases such as McMaster's work [24] [25], Jeff's method [13] and Khan's approach [23]. Also, the study shows that one of the most interesting research issues is to minimize those duplicated tests and reduce cost of executing tests. Although there are many proposed methods to resolve that issue, that issue is still remaining. Thus, it is a challenge for researchers to continuously improve the ability to remove duplicated tests.

Last, test cases are designed to reveal faults during software testing phase. The empirical studies [8], [10], [19], [20], [21], [30], [32], [39] describe that reducing test cases may impact to the ability of detect faults. Many reduction methods decrease a capability of testing and reveal those faults. Therefore, one of outstanding research challenges for researchers is to remove tests while preserving the ability to detect faults.

V. PROPOSED METHOD

For evolving software, test cases are growing dramatically. The more test cases software test engineers have, the more time and cost software test engineers consume. The literature review shows that regression testing activities consume a significant amount of time and cost. Although, a comprehensive set of regression selection techniques [8], [9], [10], [19] has been proposed to minimize time and cost, there is an available room to minimize size of tests and clean up all unnecessary test cases. Thus, removing all redundancy test cases is desirable.

There are many approaches to reduce redundancy test cases and applying an artificial intelligent concept in the test case reduction process is an innovated approach. The literature review [5], [27] shows that there are many areas of artificial intelligent concept, such as artificial neural network, fuzzy logic, learning algorithms and CBR concept. Also, it reveals that CBR has a same research issue as software testing has. The issue is that cases in the CBR system will be consistency growing bigger and larger all the time. There are four steps in CBR that can uncontrollably grow a size of the system: retrieve, reuse, revise and retain. Therefore, many CBR papers aim to reduce all redundancy cases, known as "deletion algorithms". The smaller size of CBR system is better and desirable. Due to the fact that CBR has the same problem as software testing and this paper focuses on reduction methods, therefore, this paper proposes to apply CBR deletion algorithms to the test case reduction techniques.

This paper introduces three reduction methods that apply CBR deletion algorithms: TCCF, TCIF and PCF methods.

Those techniques aim to reduce a number of test cases generated by path-oriented test case generation technique. This technique is used for white-box testing only. However, the generation methods are out of the scope of this paper.

The limitation of the proposed deletion algorithms are: (a) those methods are perfectly suitable for only white-box testing techniques and (b) path coverage may not be applicable for a large system that contains over million lines of code.

A. Example of Test Cases

Given a set of test cases generated, this study discusses the use of a number of case maintenance techniques, which have been investigated by CBR researchers in ensuring that only small amount of cases are stored in the case base, thereby reducing number of test cases should be used in software testing. Similar to what happen to software testing, a number of CBR researchers have focused on finding approaches especially for reducing cases in the CBR systems' storages.

This paper proposes to use the path coverage criteria in order to reduce redundancy test cases. This is because path coverage has a huge benefit of required very thorough testing activities. The following describes in details of the above path coverage using in the software testing field. Let $S = \{s_1, s_2, s_3, s_4, s_5\}$ to be a set of stage in the control flow graph. The control flow graph can be derived from the source-code or program. It is a white-box testing. Thus, each state represents a block of code. The techniques that aim to generate and derive test cases from the control flow graph are well-known as path-oriented test case generation techniques. These techniques are widely used to generate test cases. There are many research papers on this area. However, the test case generation techniques are out of scope in this paper.

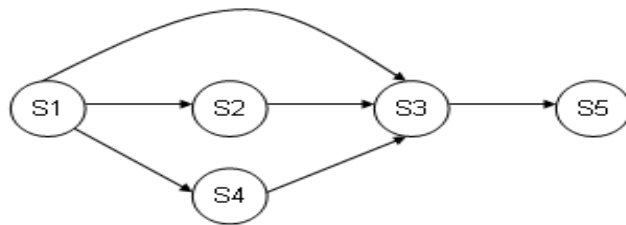


Figure 2. An Example of Control Flow Graph

From the above figure, this paper assumes that each state can reveal a fault. Thus, an ability to reveal faults of five states is equal to 5. Also, it is assumed that every single transaction must be tested. This example is used in the rest of paper.

Let $TC_n = \{s_1, s_2, \dots, s_n\}$ where TC is a test case and s_n is a stage or node in the path-oriented graph that is used to be tested. From the above figure, a set of test cases can be derived as follows:

$$TC_1 = \{s_1, s_2\}$$

$$TC_2 = \{s_1, s_3\}$$

$$TC_3 = \{s_1, s_4\}$$

$$TC_4 = \{s_1, s_2, s_3\}$$

$$TC_5 = \{s_1, s_3, s_5\}$$

$$TC_6 = \{s_1, s_4, s_3\}$$

$$TC_7 = \{s_1, s_2, s_3, s_5\}$$

$$TC_8 = \{s_1, s_4, s_3, s_5\}$$

$$TC_9 = \{s_2, s_3\}$$

$$TC_{10} = \{s_2, s_3, s_5\}$$

$$TC_{11} = \{s_3, s_5\}$$

$$TC_{12} = \{s_4, s_3\}$$

$$TC_{13} = \{s_4, s_3, s_5\}$$

From the figure 2, we assume the following: (a) each state represents a block of source code (b) each state can reveal only 1 fault; the total ability to reveal faults is 5 and (c) every single transaction in the control flow graph must be tested.

The following describes the proposed methods that apply the concept of CBR in details:

B. Path Coverage for Filtering (PCF)

Code coverage analysis is a structural testing technique (also known as white box testing). Structural testing compares test program behaviour against the apparent intention of the source code. This contrasts with functional testing (also referred to black-box testing), which compares test program behaviour against a requirements specification. Structural testing examines how the program works, taking into account possible pitfalls in the structure and logic. Functional testing examines what the program accomplishes, without regard to how it works internally. Structural testing is also called path testing since you choose test cases that cause paths to be taken through the structure of the program. The advantage of path cover is that it takes responsible for all statements as well as branches across a method. It requires very thorough testing. This is an effective substitute of other coverage criteria. The path coverage is used as coverage value in this technique. The Coverage value is combined into the addition policy for adding significant case [12]. Within the adding algorithm along with the coverage weight value stated in the review, the concept of deletion algorithm and the coverage have been proposed. The coverage value can specify how many nodes that the test case can cover. In other words, the coverage value is an indicator to measure that each test case covers nodes. It means that the higher coverage value is, the more nodes can be contained and covered in the test case.

Let $Cov(n)$ = value where Cov is a coverage value, value is a number of test cases in each coverage group and n is a coverage relationship.

The procedure of this method can be elaborated briefly as the following steps.

The first step is to determine a coverage set. From figure 2, each coverage set can be identified as follows:

$$\text{Coverage (1)} = \{TC_1\}$$

Coverage (2) = $\{TC_2\}$

Coverage (3) = $\{TC_3\}$

Coverage (4) = $\{TC_1, TC_4, TC_9\}$

Coverage (5) = $\{TC_2, TC_5, TC_{11}\}$

Coverage (6) = $\{TC_3, TC_6, TC_{12}\}$

Coverage (7) = $\{TC_1, TC_4, TC_7, TC_9, TC_{10}, TC_{11}\}$

Coverage (8) = $\{TC_3, TC_6, TC_8, TC_{11}, TC_{12}, TC_{13}\}$

Coverage (9) = $\{TC_9\}$

Coverage (10) = $\{TC_9, TC_{10}, TC_{11}\}$

Coverage (11) = $\{TC_{11}\}$

Coverage (12) = $\{TC_{12}\}$

Coverage (13) = $\{TC_{11}, TC_{12}, TC_{13}\}$

The second step is also to determine a reachability set. The reachability set can be figured out from the above coverage set, based on the given definition in this paper. Therefore, the reachability set can be identified as follows:

Reachability (TC_1) = $\{1, 4, 7\}$

Reachability (TC_2) = $\{2, 5\}$

Reachability (TC_3) = $\{3, 6, 8\}$

Reachability (TC_4) = $\{4, 7\}$

Reachability (TC_5) = $\{5\}$

Reachability (TC_6) = $\{6, 8\}$

Reachability (TC_7) = $\{7\}$

Reachability (TC_8) = $\{8\}$

Reachability (TC_9) = $\{4, 7, 9, 10\}$

Reachability (TC_{10}) = $\{7, 10\}$

Reachability (TC_{11}) = $\{5, 7, 8, 10, 11, 13\}$

Reachability (TC_{12}) = $\{6, 8, 12, 13\}$

Reachability (TC_{13}) = $\{8, 13\}$

The next step is to calculate a coverage value. This paper proposes to calculate a coverage value based on a number of test cases in each coverage group. Therefore, the coverage value can be computed as follows:

$$\begin{aligned} Cov(1) &= 1, Cov(2) = 1, Cov(3) = 1, Cov(4) = 3, Cov(5) = 3, \\ Cov(6) &= 4, Cov(7) = 6, Cov(8) = 6, Cov(9) = 1, Cov(10) = 3, \\ Cov(11) &= 1, Cov(12) = 1 \text{ and } Cov(13) = 3 \end{aligned}$$

Afterward, the step is to determine potential removable test cases. These test cases can be identified when their number of members in the reachability set is greater than a number of members in the coverage set. Therefore, the potential removal test cases are: $TC_1, TC_2, TC_3, TC_9, TC_{11}$ and TC_{12} .

The last step is to remove all test cases with minimum coverage value, in the potential removable test cases. Unfortunately, $TC_1, TC_2, TC_3, TC_9, TC_{11}$ and TC_{12} are removed due to that they have the minimum coverage value.

C. Test Case Complexity for Filtering (TCCF)

A complexity of test case is the significant criteria in this proposed method [1], [16]. In this paper, the complexity of test case measures a number of states included in each test case. We define the test case complexity as follows:

Definition 1: Let $Cplx(TC) = \{High, Medium, Low\}$ where $Cplx$ is a complexity of test case, TC is a test case and the complexity value can be measured as:

- *High* when a number of states are greater than an average number of states in the test suite.
- *Medium* when a number of states are equal to an average number of states in test suites.
- *Low* when a number of states are less than an average number of states in the test suites.

The procedures of this method can be described briefly in the following steps. The first two steps are to identify coverage and reachability set.

Next, the step is to define an auxiliary set. Test cases that can be included in the auxiliary set have a greater number of members in the reachability set than a number of members in the coverage set. From figure 2, therefore, the auxiliary set can be identified as follows:

$$\text{Auxiliary set} = \{TC_1, TC_2, TC_3, TC_9, TC_{11}, TC_{12}\}$$

Afterward, the method computes a complexity value for all test cases in the above auxiliary set. From figure 2 and test suites that contain 13 test cases, the average number of states is $(2+2+2+3+3+3+4+4+2+3+2+2+3)/13$, which is equal to 3. Based on the average number of states, the complexity value for each test case can be computed as follows:

$$\begin{aligned} Cplx(TC_1) &= Low, Cplx(TC_2) = Low, Cplx(TC_3) = Low, \\ Cplx(TC_4) &= Medium, Cplx(TC_5) = Medium, Cplx(TC_6) = \\ &Medium, Cplx(TC_9) = Low, Cplx(TC_{10}) = Medium, Cplx(TC_{11}) \\ &= Low, Cplx(TC_{12}) = Low \text{ and } Cplx(TC_{13}) = Medium \end{aligned}$$

Finally, the last step removes test cases with minimum of complexity value from the auxiliary set. Thus, $TC_1, TC_2, TC_3, TC_9, TC_{11}$ and TC_{12} are removed.

D. Test Case Impact for Filtering (TCIF)

The study [21] shows that software is error-ridden in part because of its growing complexity. Software is growing more complex every day. The size of software products is no longer measured in thousands of lines of code, but it measures in millions. Software developers already spend approximately 80 percent of development costs [18] on identifying and correcting defects, and yet few products of any type other than software are shipped with such high levels of errors. Other factors contributing to quality problems include marketing strategies, limited liability by software vendors, and decreasing returns on testing and debugging, according to the study. At the core of these issues is difficulty in defining and measuring software quality. Due to the fact that defining and measuring a quality of software is important and difficult, the impact of inadequate

testing must not be ignorance. The impact of inadequate testing could be lead to the problem of poor quality, expensive costs and huge time-to-market. In conclusion, software testing engineers require identifying the impact of each test case in order to acknowledge and understand clearly the impact of ignoring some test cases.

In this paper, an impact value is an impact of test cases in term of the ability to detect faults if those test cases are removed and not be tested. We define the test case impact as follows:

Definition 2: Let $Imp(TC) = \{High, Medium, Low\}$ where Imp is an impact if a test case is removed, TC is a test case and the impact value can be measured as:

- *High* when the test case has revealed at least one fault for many times.
- *Medium* when the test case has revealed faults for only one time.
- *Low* when the test case has never revealed faults.

The procedure of this method is similar to the previous method. The only different is that this method aims to use an impact value instead of complexity value. Therefore, the fire three steps are to: identify coverage set, define reachability set and determine an auxiliary set. Afterward, the next step is to compute and assign an impact value. The method computes the impact value for all test cases in the above auxiliary set. From figure 2, the impact value for each test case can be computed as follows:

$Imp(TC_1) = Low$, $Imp(TC_2) = High$, $Imp(TC_3) = Medium$, $Imp(TC_4) = Low$, $Imp(TC_5) = High$, $Imp(TC_6) = Medium$, $Imp(TC_9) = Low$, $Imp(TC_{10}) = Low$, $Imp(TC_{11}) = Low$, $Imp(TC_{12}) = Low$ and $Imp(TC_{13}) = Low$

Finally, the last step removes test cases with minimum of impact value from the auxiliary set. Thus, TC_1 , TC_4 , TC_7 , TC_9 , TC_{10} , TC_{11} , TC_{12} and TC_{13} are removed.

VI. EVALUATION

This section describes an experiments design, measurement metrics and results.

This paragraph designs an experiment used to evaluate and determine the best reduction methods. This paper proposes the following three steps. First, the experiment proposes to randomly generate 2,000 test data used in the telecommunication industry. In this experiment, the test data is represented as test case. Second, the experiment executes reduction methods with the generated test cases and compares among the following reduction methods: RD, UD, FD, FUD, ICF and three proposed methods (e.g. TCCF, TCIF and PCF). This step randomly simulates defects for each test case in order to determine an ability to reveal faults. Third, the experiment aims to run the above methods for 10 times in order to calculate the average value for each metric. The metrics used in this experiment are described in details in next section. Afterward, the experiment compares the values and

evaluates a result by generating a comparison graph in order to determine the most recommended reduction approach.

The following table lists the description of each test data that need to be generated randomly.

TABLE II. A FORM OF TEST CASES

Attribute	Description	Data Type
Test Id	A unique index to reference test data. The value is a sequence number, starting at 1.	Numeric
A Set of Input Data		
Full Name	A first and last name who own the mobile phone.	String
Name	A mobile brand name. The value is a range of iPhone, BlackBerry, Nokia, LG, Sony Ericsson and Samsung.	String
Model	A mobile model.	String
Price	A price of mobile. The unit of price is baht.	Numeric
Weight	A weight of mobile. The unit of weight is gram (g).	Numeric
Height	A height of mobile. The unit of height is centimeter (cm).	Numeric
Graphics	A graphics mode option. The value can be T or F	Boolean
WAP	A WAP mode. The value is T or F	Boolean
Color	A mobile color. The color can be Black, Gold, Silver, Blue, and White	String
Game	A game mode. The value is T or F	Boolean
Warranty	A mobile warranty. The unit of warranty is month.	Numeric

The following table describes an approach to generate random data using the above attributes respectively.

TABLE III. APPROACH TO GENERATE RANDOM TEST CASES

Attribute	Approach
Test Id	Generate randomly from the following combination: $t + Sequence\ Number$. For example, t_1 , t_2 , t_3 , ..., t_n .
Name	Random from the following values: iPhone, BlackBerry, Nokia, LG, Sony and Samsung.
Model	Random from the following values: iPhone – iPhone 2G and iPhone 3G. BlackBerry – BlackBerry Bold 9000, BlackBerry Bold 9700, BlackBerry

	Curve 8300, BlackBerry Curve 8520, BlackBerry Curve 8900, BlackBerry Pearl, BlackBerry Pearl Flip 8200 and BlackBerry Storm. Nokia – Nokia N97 Mini, Nokia E90, Nokia E72, Nokia 8800, Nokia N86 and Nokia E66. LG – Lotus Elite LX610, Accolade VX5600, Chocolate Touch VX8575, Arena GT950 and eXpo GW820. Sony Ericsson – Xperia X10, Vivaz Pro, Xperia X10 Mini Pro, Elm and Aspen. Samsung – Samsung Gravity 2, Behold II, Comeback, SGH-T139, SGH-T659 and SGH-T239.
Price	Random from the following values: High: 30,000 baht – 50,000 baht Medium: 10,000 baht – 29,999 baht Low: 3,000 baht – 9,999 baht
Weight	Random from the following values: 4-15
Height	Random from the following values: 5-15
Graphics	Random from the following values: True or False
WAP	Random from the following values: True or False
Color	Random from the following values: Black, Gold, Silver, Blue and White
Game	Random from the following values: True or False
Warranty	Random from the following values: Long: 12 months – 18 months Medium: 6 months – 11 months Short: 1 month – 5 months

The paragraph lists the measurement metrics used in the experiment. The first measurement is a number of test cases. The large number of test cases consumes time, effort and cost more than the smaller size of test cases. Many reduction or minimization approaches [6], [7], [8], [9], [10], [11], [19], [24], [25], [36], [37], [39] have been proposed to minimize size of test cases. This has proven that size is one of important metrics in software testing area. The second is an ability to reveal faults. It aims to measure the percentage of faults detection. One of the goals of test case with a set of data is to find defects. Thus, this metric is important criteria to measure and determine which reduction methods can preserve the high ability to reveal faults. The last measurement is a total of reduction time: It is the total number of times running the reduction methods in the experiment. This metric is related to time used during execution time and maintenance time of test case reduction methods. Therefore, less time is desirable.

This paragraph discusses an evaluation result of the above experiment. This section presents the reduction methods results in term of: (a) a number of test cases (b) ability to reveal faults and (c) total reduction time. The comparative

methods are: RD, UD, FD, FUD, ICF, TCCF, TCIF and PCF. Additionally, this section shows a graph format. There are two dimensions in the following graph: (a) horizontal and (b) vertical axis. The horizontal represents three measurements whereas the vertical axis represents the percentage value.

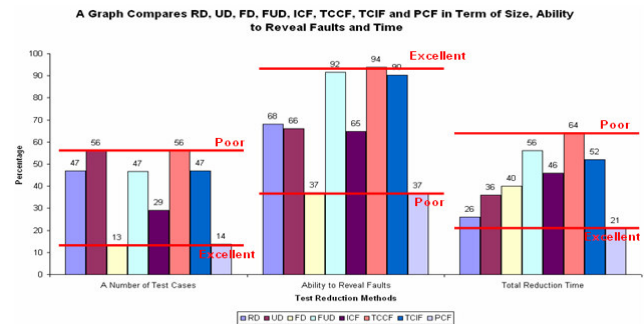


Figure 2. A Graph Comparison of Deletion Methods

The above graph presents that both of FD and PCF minimize a number of test cases by far better than other reductions methods, approximately over 15%. Meanwhile, both of them are the worst methods for preserving an ability to reveal faults. FUD, TCCF and TCIF are best top three methods to reserve a capability to detect faults. They are greater than other methods over 22%. Unfortunately, they are also the worst three methods that require a lot of time during a reduction process. In the mean time, both of RD and PCF take the least total reduction time among other methods.

From the above figure, this study determines and ranks the above comparative methods into five ranking: 5-Excellent, 4-Very good, 3-Good, 2-Normal and 1-Poor. This study uses a maximum and minimum value to find an interval value for ranking those methods.

For a number of test cases, the maximum and minimum percentage is 56% and 13%. The different between maximum and minimum value is 43%. An interval value is equal to a result of dividing the different values by 5. As a result, the interval value is 8.6. Thus, it can be determined as follows: 5-Excellent (since 13% to 21.6%), 4-Very good (between 21.6% and 30.2%), 3-Good (between 30.2% and 38.8%), 2-Normal (between 38.8% and 47.4%) and 1-Poor (from 47.4% to 56%).

For an ability to reveal faults, the maximum and minimum percentage is 94% and 37%. The different value is 57%. The interval value is 11.4. Therefore, it can be determined as follows: 5-Excellent (since 82.6% to 94%), 4-Very good (between 71.2% and 82.6%), 3-Good (between 59.8% and 71.2%), 2-Normal (between 48.4% and 59.8%) and 1-Poor (from 37% to 48.4%).

For a total reduction time, the maximum and minimum percentage is 64% and 21%. The different between maximum and minimum value is 43%. An interval value is equal to a result of dividing the different values by 5. As a result, the interval value is 8.6. Thus, it can be determined as follows: 5-Excellent (since 21% to 29.6%), 4-Very good (between 29.6%

and 38.2%), 3-Good (between 38.2% and 46.8%), 2-Normal (between 46.8% and 55.4%) and 1-Poor (from 55.4% to 64%).

Therefore, the experiment result of those eight comparative methods can be shown below:

TABLE IV. A COMPARISON OF TEST CASE REDUCTION METHODS

Algorithm	A Number of Test Cases / Size	Abili ty to Reveal Faults	Total Reduction Time
Random Deletion (RD)	2	3	5
Utility Deletion (UD)	1	3	4
Footprint Deletion (FD)	5	1	3
Footprint Utility Deletion (FUD)	2	5	1
Iterative Case Base Filtering (ICF)	4	3	3
Test Case Complexity for Filtering (TCCF)	1	5	1
Test Case Impact for Filtering (TCIF)	2	5	2
Path Coverage for Filtering (PCF)	5	1	5

The above result suggests that FD and PCF is perfectly suitable for a scenario that does not directly concern about an ability to reveal faults and total reduction time. Both of FD and PCF are two of the most excellent methods to minimize a number of test cases. Meanwhile, FUD, TCCF and TCIF are the most recommended methods to delete tests while preserving the ability to detect faults. In addition, both of RD and PCF are excellent in case that total reduction time is matter.

The above graph presents that both of FD and PCF minimize a number of test cases by far better than other reductions methods, approximately over 15%. Meanwhile, both of them are the worst methods for preserving an ability to reveal faults. FUD, TCCF and TCIF are best top three methods to reserve a capability to detect faults. They are greater than other methods over 22%. Unfortunately, they are also the worst three methods that require a lot of time during a reduction process. In the mean time, both of RD and PCF take the least total reduction time among other methods.

The evaluation result suggests that FD and PCF is perfectly suitable for a scenario that does not directly concern about an ability to reveal faults and total reduction time. Both of FD and PCF are two of the most excellent methods to minimize a number of test cases. Meanwhile, FUD, TCCF and TCIF are the most recommended methods to delete tests while preserving the ability to detect faults. In addition, both of RD and PCF are excellent in case that total time is matter.

VII. CONCLUSION

This paper reveals that there are many research challenges and gaps in the test case reduction area. Those challenges and gaps can give the research direction in this field. However, the research issues that motivated this study are: a large number of test cases and a decrease of ability to reveal faults. This paper combines an approach to maintain CBR and test case reduction in order to minimize a number of redundancy tests while maintaining an ability to detect faults.

In this paper, we proposed deletion algorithms to reduce a number of test cases that generated by widely-used white-box testing techniques. As part of our research, we conducted an experiment with 2,000 test cases used in the telecommunication industry in Thailand. Consequently, our evaluation result reveal that the proposed method is one of the most recommendation techniques to maintain ability to reveal faults and minimize a number of redundancy test cases. However, the limitation of the proposed techniques is that path coverage may be not an effective coverage factor for a huge system that contains million lines of code. This is because it requires an exhaustive time and cost of identify coverage from a huge amount of codes.

In future research, we plan to develop deletion algorithms with other coverage factors, such as function coverage and block-statement coverage. Also, we aim to implement deletion algorithms for multiple test suites. Finally, the evaluation experiment should be conducted for a large commercial system

REFERENCES

- [1] Barry Smyth and Keane. "Remembering To Forget: A Competence Preserving Deletion Policy for Case-Based Reasoning Systems." Proceedings of the 14th International Joint Conference on Artificial Intelligence. Montréal, Québec, Canada: Morgan-Kaufman Inc., 1995. 377-382.
- [2] Beizer, Boris. Software Testing Techniques. New York, USA: Van Nostrand Reinhold Inc., 1990.
- [3] BO Qu, Changhai Nie, Baowen Xu and Xiaofang Zhang. "Test Case Prioritization for Black Box Testing." Proceeding with 31st Annual International Computer Software and Applications Conference (COMPSAC 2007). Beijing, China, 2007. 465-474.
- [4] Boehm, B.W. "A spiral model of software development and enhancement." IEEE Software Engineering (IEEE Computer Society), 1988: 61-72.
- [5] Daengdej, Jirapun. Adaptable Case Base Reasoning Techniques for Dealing with Highly Noise Cases. PhD Thesis, The University of New England, Australia: The University of New England, 1998.
- [6] Gregg Rothermel and Mary Jean Harrold. "A Safe, Efficient Regression Test Selection Technique." ACM Transactions on Softw. Eng. And Methodology, 1997: 173-210.
- [7] Gregg Rothermel and Mary Jean Harrold. "Analyzing Regression Test Selection Techniques." IEEE Transactions on Software Engineering, 1996: 529-511.
- [8] Gregg Rothermel, Mary Jean Harrold, Jeffery Ostrin and Christie Hong. "An Empirical Study of the Effects of Minimization on the Fault Detection Capabilities of Test Suites." Proceedings of IEEE International Test Conference on Software Maintenance (ITCSM'98). Bethesda, Maryland, USA: IEEE Computer Society, 1998. 33-43.
- [9] Gregg Rothermel, Mary Jean Harrold, Jeffery von Ronne and Christie Hong. "Empirical Studies of Test-Suite Reduction."

- Journal of Software Testing, Verification, and Reliability 12, no. 4 (December 2002): 219-249.
- [10] Gregg Rothermel, Roland H. Untch, Chengyun Chu and Mary Jean Harrold. "Prioritizing Test Cases For Regression Testing." IEEE Transactions on Software Engineering, 2001: 929-948.
- [11] Jefferson Offutt, Jie Pan and Jeffery M. Voas. "Procedures for Reducing the Size of Coverage-based Test Sets." Proceedings of the Twelfth International Conference on Testing Computer Software. Washington D.C, USA, 1995. 111-123.
- [12] Jun Zhu and Quiang Yang. "Remembering To Add Competence-preserving Case Addition Policies for Case Base Maintenance." Proceedings of the 16th International Joint Conference in Artificial Intelligence. Stockholm, Sweden : Morgan Kaufmann Publishers Inc, 1999. 234-241.
- [13] Kaner, Cem. "Exploratory Testing." Quality Assurance Institute Worldwide Annual Software Testing Conference. Orlando, Florida, USA: Florida Institute of Technology, 2006.
- [14] Lehmann, E. and J. Wegener. "Test case design by means of the CTE XL." Proceedings of the 8th European International Conference on Software Testing,. Copenhagen, Denmark: ACM Press, 2000. 1-10.
- [15] Mary Jean Harrold, Rajiv Gupta and Mary Lou Soffa. "A Methodology for Controlling the Size of A Test Suite." ACM Transactions on Software Engineering (ACM) 2, no. 3 (July 1993): 270-285.
- [16] Nicha Kosindrdecha and Jirapun Daengdej. A Deletion Algorithm for Case-Based Maintenance Based on Accuracy and Competence. MS Thesis, Faculty of Science and Technology, Assumption University, Bangkok, Thailand: Assumption University, 2003.
- [17] Nicha Kosindrdecha and Siripong Roongruangsuwan. "Reducing Test Cases Created by Path Oriented Test Case Generation." Proceedings of the AIAA Conference and Exhibition. Rohnert Park, California, USA: NASA AIAA, 2007.
- [18] NIST. The economic impacts of inadequate infrastructure for software testing. Technical Report, USA: NIST, 2002.
- [19] Rothermel, G., R.H. Untch, C. Chu and M.J. Harrold. "Test case prioritization: An empirical study." Proceedings of the 15th IEEE International. Oxford, England, UK: IEEE Computer Society, 1999. 179-188.
- [20] S. Elbaum, A. G. Malishevsky and G. Rothermel. "Prioritizing Test Cases for Regression Testing." Proceedings of the International Symposium on Software Testing and Analysis. 2000. 102-112.
- [21] S. Elbaum, A. Malishevsky, and G. Rothermel. "Test Case Prioritization: A Family of Empirical Studies." IEEE Trans. on Software Engineering 28 (February 2002): 159-182.
- [22] S. Elbaum, P. Kallakuri, A. G. Malishevsky, G. Rothermel, and S. Kanduri. "Understanding the effects of changes on the cost-effectiveness of regression testing techniques." Journal of Software Testing, Verification, and Reliability 13, no. 2 (June 2003): 65-83.
- [23] Saif-ur-Rebman Khan and Aamer Nadeem. "TestFilter: A Statement-Coverage Based Test Case Reduction Technique." Proceedings of 10th IEEE International Multitopic Conference. Islamabad, Pakistan, 2006.
- [24] Scott McMaster and Atif Memon. "Call Stack Coverage for GUI Test-Suite Reduction." Proceedings of the 17th IEEE International Symposium on Software Reliability Engineering (ISSRE 2006). North Carolina, USA, 2006.
- [25] —. "Call Stack Coverage for Test Suite Reduction." Proceedings of the 21st IEEE International Conference on Software Maintenance (ICSM'05). Budapest, Hungary, 2005. 539-548.
- [26] —. "Fault Detection Probability Analysis for Coverage-Based Test Suite Reduction." Proceedings of IEEE International Conference on Software Maintenance (ICSM 2007). Paris, France, 2007. 335-344.
- [27] Siripong Roongruangsuwan and Jirapun Daengdej. Techniques for improving case-based maintenance. MS Thesis, Faculty of Science and Technology, Assumption University, Bangkok, Thailand: Assumption University, 2003.
- [28] Siripong Roongruangsuwan and Jirapun Daengdej. Test case reduction. Technical Report 25521, Bangkok, Thailand: Assumption University, 2009.
- [29] Smyth, Barry. Case Based Design. PhD Thesis, Department of Computer Science, Trinity College, Dublin, Ireland: Trinity College, 1996.
- [30] Sprenkle, S., S. Sampath and A. Souter. "An empirical comparison of test suite reduction techniques for user-session-based testing of web applications." Journal of Software Test. Verificat. Reliabil., 2002: 587-596.
- [31] Sreedevi Sampath, Sara Sprenkle, Emily Gibson and Lori Pollock. "Web Application Testing with Customized Test Requirements – An Experimental Comparison Study." Proceedings of the 17th International Symposium on Software Reliability Engineering (ISSRE'06). Raleigh, NC, USA: IEEE Computer Society, 2006. 266 - 278.
- [32] Todd L. Graves, Mary Jean Harrold, Jung-Min Kim, Adam Porter and Gregg Rothermel. "An Empirical Study of Regression Test Selection Techniques." ACM Transactions on Software Engineering and Methodology (TOSEM), 2001: 184-208.
- [33] W. Eric Wong, J. R. Horgan, Saul London and Hira Agrawal. "A Study of Effective Regression Testing in Practice." Proceedings of 8th IEEE International Symposium on Software Reliability Engineering (ISSRE'97). California, USA: IEEE Computer Society, 1997. 264.
- [34] W. Eric Wong, Joseph R. Horgan, Saul London and Aditya P. Mathur. "Effect of Test Set Minimization on the Fault Detection Effectiveness of the All-Uses Criterion." Proceedings of the 17th International Conference on Software Engineering. Seattle, USA: ACM, 1995. 41-50.
- [35] Wilson, David C. A Case-Based Maintenance: The husbandry of experiences. PhD Thesis, Computer and Science, Indiana University, USA: Indiana University, 2001.
- [36] Xiaofang Zhang, Baowen Xu, Changhai Nie and Liang Shi. "An Approach for Optimizing Test Suite Based on Testing Requirement Reduction." Journal of Software (in Chinese), 2007: 821-831.
- [37] Xiaofang Zhang, Baowen Xu, Changhai Nie and Liang Shi. "Test Suite Optimization Based on Testing Requirements Reduction." International Journal of Electronics & Computer Science 7, no. 1 (2005): 9-15.
- [38] Xue-ying MA, Bin-kui Sheng, Zhen-feng HE and Cheng-qing YE. "A Genetic Algorithm for Test-Suite Reduction." Proceedings of IEEE International Conference on Systems, Man and Cybernetics. Hangzhou, China: ACM Press, 2005. 133-139.
- [39] Yanbing Yu, James A. Jones and Mary Jean Harrold. "An Empirical Study of the Effects of Test-Suite Reduction on Fault Localization." Proceedings of International Conference on Software Engineer (ICSE'08). Leipzig, Germany: ACM, 2008. 201-210.

Nat Traversal for Video Streaming Applications

Omar A. Ibraheem
National Advanced IPv6 Centre of
Excellence (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA

Omer Amer Abouabdalla
National Advanced IPv6 Centre
Excellence (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA

Sureswaran Ramadass
National Advanced IPv6 Centre
Excellence (NAv6)
Universiti Sains Malaysia
11800, Penang, MALAYSIA

Abstract— This paper presents a novel method that exploits the strength features of two streaming protocols (Real-time Transport Protocol (RTP) and Hypertext Transfer Protocol (HTTP)) to overcome the Network Address Translation (NAT) and firewall traversal problem. The proposed solution is able to bypass the RTP over all kinds of NATs (including symmetric NATs) by adding extra fields to the RTP/UDP packet at transport layer in the sender side. The NAT and firewall will detect these packets as TCP packets on the channel that initialized the connection. The receiver side will then remove the extra fields and recover the packets to their original content. The proposed work involves adding two modules, one at the client and the other at the video streaming server. The proposed work also avoids any modification to the NAT or the RTP protocol itself.

Keywords- NAT Traversal; Video Streaming; RTP; TCP; UDP; Windows OS.

I. INTRODUCTION

Video streaming is considering one of the famous technologies which is used today. It provides the ability to playback video files remotely through computer networks. The demand for this technology is rapidly increasing due to wide spread of Internet and increasing of the network bandwidth[1]

There are two main application layer protocols that are used for video streaming: RTP and HTTP. Although the RTP protocol is originally developed for video streaming and the HTTP protocol is originally developed for browsing, which has many weakness points when dealing with video streaming, HTTP protocol is still used and more spread than RTP when used with video streaming. This is due to the simplicity of using the HTTP protocol and in order to avoid the problems that faced the RTP protocol.

While HTTP protocol uses one TCP port at the transport layer, RTP can use many ports. RTP can use UDPs or TCPs ports at the transport layer depending on how much the packet path is suffered from packet loss [2]. In low packets loss environment, the use of RTP over UDP protocol is preferable, since in media streaming, the small ratio of packets loss better than packets delay. Hence, the higher reliability of the TCP is not desired[3]. UDP/RTP has also the multicasting feature and has the ability to deal with real time communication due to its features in bandwidth, jitter, reliability and end node's processing.

RTP/TCP can cause the video streaming to suffer from discontinuity because the need to reordering, acknowledgement, and retransmission and the packets, whereas RTP/UDP can suffer from dropping the packets by some filters (firewalls) in the Internet Service Provider (ISPs). Some ISPs drop UDP packets because they are connectionless hence unfair against TCP traffic. They also need high processing power and memory to ensure security [4]. But the main issue that can occur is when using the RTP with the Network Address Translation (NAT). NAT drops any RTP/UDP or RTP/TCP packets that are initialized from the outside (Internet) when incoming to the end-systems (behind the NAT).

The NAT is a technology that permits many computers on the same network to share a public Internet Protocol (IP) address for accessing the Internet. The main reason behind the wide spread of using the NAT is the limited number of the available IPv4 addresses [5].

The use of RTP/UDP or RTP/TCP video streaming is started with a TCP connection that is established by a request from the client to the server, after initial negotiation using the RTSP protocol on the same established TCP channel, the server starts video streaming through UDP or TCP ports initialized from the server not through the established RTSP/TCP channel [2].

The NAT permits to pass the outgoing connections requests produced from a host behind the NAT into the outside network (like Internet) [6], however it does not permit to pass any connection request produced from the outside network (like Internet) to any host behind the NAT [7]. This is because the translation table entry is constructed only when a client (behind the NAT) initializes a request to connect to a host on the outside network (Internet) [8], [9]. If the initialized request comes from a host outside the network of the NAT into the inside network, the NAT cannot identify the destination host for this request and the connection between the outside host and the inside one cannot be occur [8], [10]. Regarding to the RTP/UDP video streaming, the NAT will not allow the UDP video streaming channels to pass to the client behind the NAT, since the RTP/UDP channels are initially established from the server (on the Internet).

Considering the RTP weakness points, the HTTP protocol, is the preferable choice for video streaming. However, HTTP protocol also has known weakness points: the user can suffer

from quality reduction and playback discontinuity due to the probing behaviour of TCP protocol. This can also cause an oscillating throughput and slow recovery of the packet rate.

In contrast, the UDP protocol provides a mean to keep the desired sending rate constant. It also keeps streaming smooth and eliminates the TCP related processing.

This paper presents a novel method to utilize the benefits of both TCP and UDP. The proposed method enables the video streaming to traverse the NAT by converting each RTP/UDP and RTCP/UDP packet into fake TCP packet just before being sent (at data link layer) by adding a fabricated TCP header before each UDP video streaming packet and making the necessary modifications to the length and checksums fields.

These fabricated TCP packets will pass the NAT, since they will be transmitted on the channel (IP, TCP port) that firstly initialized (RTSP/TCP channel) by the client behind the NAT. In this paper, this channel is called the *active channel*.

The receiver, on the other side has to restore the original UDP packet before being processed by the corresponding transport layer. The restoration is based on a specific signature. In order to restore the packets, every fabricated TCP packet has to have a known signature. Depending on that signature, the receiver will restore the original packet. All of the previous changes are performed at the data link layer.

The rest of this paper is organized as follows: section II, looks at some related work. In section III, the proposed methodology and algorithm are presented. In section IV, the experiments of the implemented proposed method and its discussions are described. In section V, the evaluation of the proposed method and comparisons between the proposed method and the existing technologies are presented. The paper is concluded in section VI.

II. RELATED WORK

Limited to our knowledge, no many similar works are presented. However, [4] present a method to overcome the RTP/UDP issues by putting a proxy server between the client and the streaming server (at the global network). The proxy receives a HTTP request (TCP) from the client and translates it to a RTSP/RTP request to the server (TCP+UDP). The proxy has two different connections (one for the client and the other for the streaming serve). The main function of the proxy is to translate the HTTP streaming protocol into RTSP/RTP streaming protocol. This can overcome the NAT problem due to that the HTTP request (TCP) is initialized by the client and the reply will pass through the same TCP port. However a third device is needed. In addition it is still using the constraints of the TCP between the proxy and the client (e.g. retransmission and reordering ...etc) (in addition to the increase of traffic to the network). Another issue is that there are too many operations in order to convert a complete application protocol into another one. Beside, this method loses the real time property that is needed for end to end communication because all the packets must be forwarded at the proxy server.

III. PROPOSED METHODOLOGY

In this work, both the client and the server are assumed to convert all the RTP/UDP streaming packets into fabricated TCP packets that can be sent to the other side using the *active channel*.

This fabrication process which is implemented for Windows Operating System (OS) requires a full control of the incoming/outgoing packets. However, there is the issue of source code of the TCP/IP (non open source for Windows OS) is not readily accessible and Windows does not allow the manipulation of the packets in any TCP/IP protocol suite from level above the TCP/IP driver layer.

To overcome the inaccessibility issue, a hooking technique is used in order to control the (frame/packet) at the point that links between the protocol driver and the NIC card(s), which is represented by the Network Driver Interface Specification (NDIS).

Hooking is a technique that can convert the calling of one operating system function into a new one that in turn calls the old one. The new function can do extra job before moving the execution to the old one. This can be done without the need for the source code of the old one [11].

The proposed modules is implemented and run in windows user mode. When the module can hook the NDIS, it can monitor, control, add, and modify the NDIS incoming/outgoing packets easily.

The NDIS-Hooking driver inserts itself between TCP/IP and all of the adapters that bind with it as shown in figure (1).

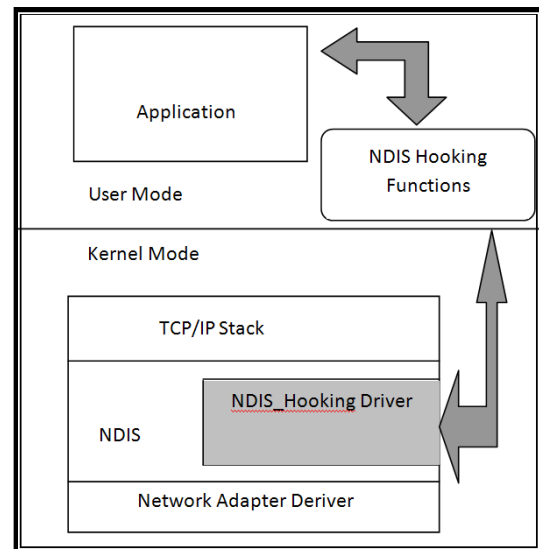


Figure 1. NDIS hooking driver with relation to user mode¹

When TCP/IP sends a packet, it reaches the NDIS-Hooking driver (as a frame) before sending to the adapter. Likewise, packets that are to be indicated (received) on TCP/IP will go to the NDIS-Hooking driver first.

¹ <http://www.ntkernel.com/w&p.php?id=7>

The fabricated TCP header is inserted/deleted in the data link frame, this means that the original RTP/UDP protocol is used without modification. Nonetheless the fabricated packets can still bypass the NAT as authenticated ones.

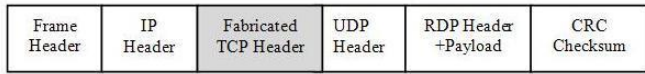


Figure 2. Proposed frame structure

As these extra bytes (fabricated TCP header) will be added when the packet is in the data link layer, this may cause the packet to exceed the Maximum Transfer Unit (MTU) of the network. Since, no packet must exceed the Maximum Transfer Unit (MTU) of the network [12], [13], therefore, the sender's MTU must be decreased by length of the fabricated TCP header length (20 bytes).

The whole proposed system is composed of two main modules. The first module resides on the streaming client while the second resides on the streaming server. Figure (3) shows the video streaming network topology.

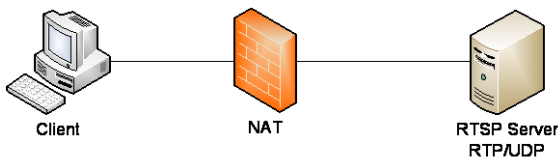


Figure 3. Video streaming network topology

Each module consists of the following components:

A component (hooking function in Fig. 1) that provides a way to access the frame at the data link layer. This component accesses the frames in data link layer which is in the kernel mode and moves it into the user mode and vice versa.

A component that finds the required frame based on its content. This component extracts the specified packets from the frames which have to be changed (fabricated/restored) depending on sending direction (income/outcome).

A component that makes the required modifications (fabricating/restoring) to the predetermined packets. This component changes the predetermined packets depending the sending direction (send/receive). In sending, the component changes the RTP/UDP packet into fabricated TCP packet. In receiving, the component restores the fabricated TCP packet into its original RTP/UDP content. This component also recomputes length and checksums.

A. Client Side Module

As mentioned earlier, the module has to access the kernel (at data link layer). This is done by accessing the NDIS driver. The module listens until a packet event has occurred. There are two possible scenarios:

Incoming packet: If the packet is coming from the streaming server, then the program will look for the TCP that contains an RTSP packet. If this RTSP packet contains both

the client's and server's streaming ports, then record this connection's information into an array. This is happened normally at the setup phase of the RTSP connection. Later (when the video streaming data is transferred), the client will check every TCP packet if it contains a specified signature. If this signature is raised (in the TCP header), this mean that this TCP packet is fabricated and it contains the original RTP/UDP packet. The program will remove the TCP header and recomputed the UDP and IP checksums. All these steps are done before sending the packet to the rest of TCP/IP protocol stack

Outgoing packet: If the packet is outgoing to the streaming server and the outgoing packet was a RTP/UDP packet, then insert a new fabricated TCP header before the UDP header. This fabricated TCP header contains the TCP connection information taken from the appropriate record from an array containing all streaming connections' details. This TCP header also contains a specified signature that has to be recognized from the streaming server in order to return the packet back to its original RTP/UDP packet. This operation also needs to recompute the checksums. All these steps are done before sending the packet to the adapter. Figure 4 shows the flowchart of client side module.

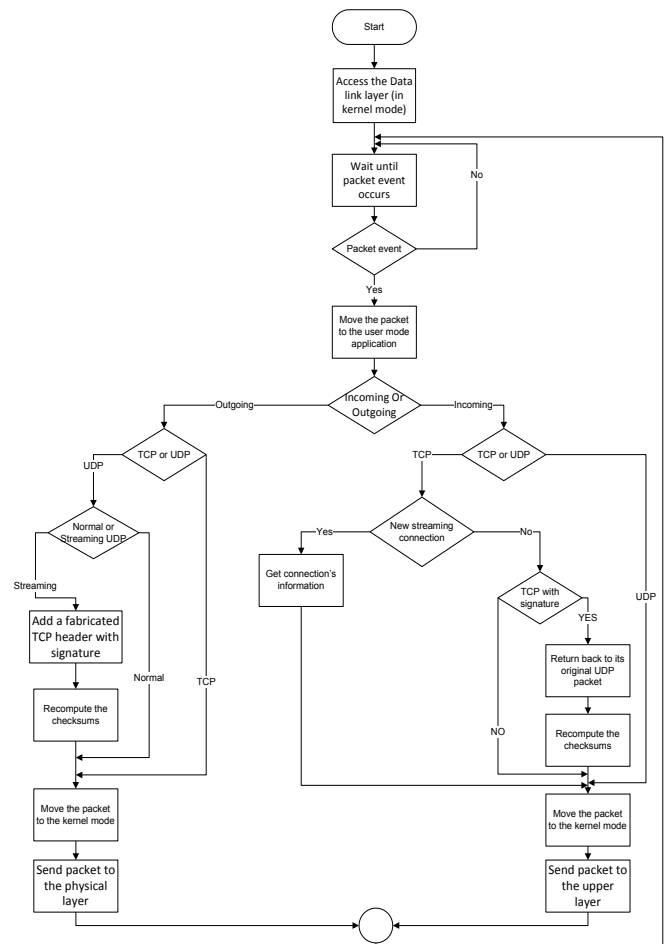


Figure 4. Flowchart of the client side module

B. Server Side Module

In server side module, similar steps to the client are also implemented. The difference is that the system gets the RTSP connection's details from the outgoing TCP packet instead of incoming TCP packet in the client. Figure (5) shows the flowchart of the main steps of the server module.

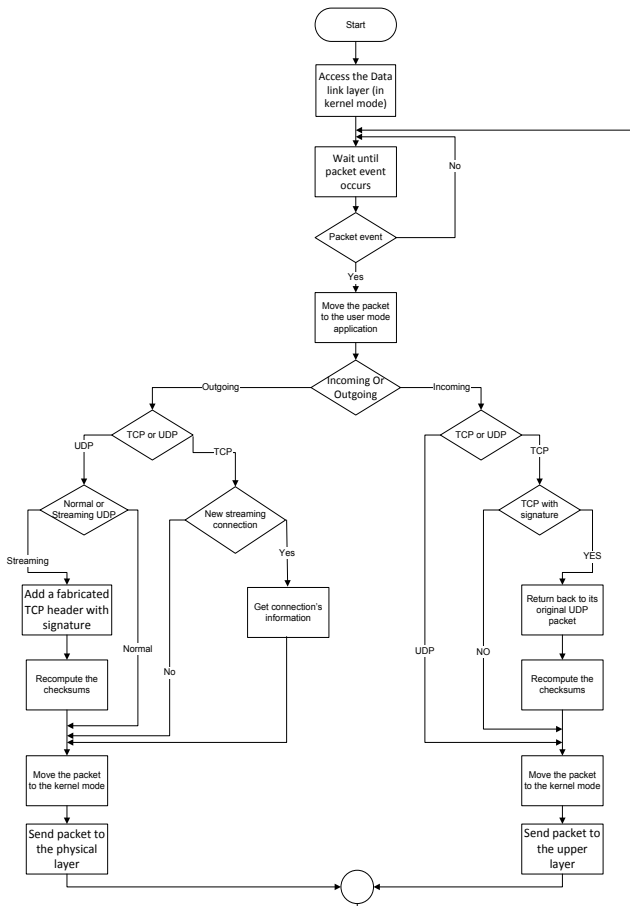


Figure 5. Flowchart of the server side module

IV. EXPERIMENTS AND DISCUSSIONS

A. Experiments Setup

In this experiment, we use three PCs running windows XP. Two PCs with one LAN card (client and the server). The other PC (working as a NAT) contains two LAN cards.

RedMug streaming server commercial software is used on the server site. The VLC media player (version 1.0.5) is used on the client side. The VLC media player is set to use the RTSP protocol by giving a URL of one movie on the streaming server. The proposed method (client and server modules) is implemented in VC++.Net Framework and it is running in windows OS environment in user mode. A windows device driver (Windows Packet Filter Kit "winpkfilter 3.0" from NT Kernel Resources, <http://www.ntkernel.com>) is used for the hooking purpose.

B. Experimental Results and Discussion

In the first experiment (before using the proposed method), the client tries to access the movie on the streaming server using the above system configuration. The connection's establishment and the video streaming negotiations between the client and the server are established normally. However, the connection fails at the stage of data streaming transformation (see Fig. 6).

```
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 5. Received RTSP message from 150.150.100.110:62452. Message: 'DESCRIBE rtsp://150.150.100.110:62452/sample.mpg'
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 5. Is the 'C:\redMUG\Contents\sample.mpg' MP4 file in cache? 'true'
RTSPSession identifier: 5. Video track name to stream: trackID/2, audio track name to stream: trackID/1
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 5. Received RTSP message from 150.150.100.110:62452. Message: 'SETUP rtsp://150.150.100.110:62452/sample.mpg'
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
Session: 257655783;timeout=360
RTSPSession identifier: 5. Received RTSP message from 150.150.100.110:62452. Message: 'SETUP rtsp://150.150.100.110:62452/sample.mpg'
Session: 257655783
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 5. Received RTSP message from 150.150.100.110:62452. Message: 'PLAY rtsp://150.150.100.110:62452/sample.mpg'
Session: 257655783
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 5. Called the seek method. Start time in secs: 0.000000.
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
Session: 257655783;timeout=360
Client 6 is opened
RTSPSession identifier: 5. Received RTSP message from 150.150.100.110:62452. Message: 'TEARDOWN rtsp://150.150.100.110:62452/sample.mpg'
Session: 257655783
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 5. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 5. Connection from 150.150.100.110:62452 closed.
RTSPSession identifier: 6. Connection arrived from 150.150.100.110:62453.
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 6. Received RTSP message from 150.150.100.110:62453. Message: 'OPTIONS rtsp://150.150.100.110:62453/sample.mpg'
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 6. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 6. Received RTSP message from 150.150.100.110:62453. Message: 'DESCRIBE rtsp://150.150.100.110:62453/sample.mpg'
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 6. Is the 'C:\redMUG\Contents\sample.mpg' MP4 file in cache? 'true'
RTSPSession identifier: 6. Video track name to stream: trackID/2, audio track name to stream: trackID/1
RTSPSession identifier: 6. Sent to client: 'RTSP/1.0 200 OK'
RTSPSession identifier: 6. Received RTSP message from 150.150.100.110:62453. Message: 'SETUP rtsp://150.150.100.110:62453/sample.mpg'
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 6. Sent to client: 'RTSP/1.0 500 Internal Server Error'
RTSPSession identifier: 6. Connection from 150.150.100.110:62453 closed.
```

Figure 6. Connection breakdown when data streaming transforming began (server side)

The reason for the success of the initialization of the client-server connection and all the negotiations needed to transfer the video streaming are that the connection request is a TCP and the initialization is coming from the client (behind the NAT) and the video streaming negotiations are done by the RTSP that uses the *active channel*. However, the client could not receive the video streaming data since the NAT dropped the RTP/UDP video streaming packets. The client then sends a teardown command to inform the server that the negotiation is over. The client starts one additional negotiation tries before it close the connection.

In the second experiment, we used the proposed client and server modules. After running, the two modules start monitoring the data link frames. The client monitors the outgoing streaming request while the server monitors the incoming streaming request.

When the client request a video streaming from the server, The connection's establishment and the video streaming negotiations between the client and the server are established normally and the client started to display the video streaming data as shown in figure (7A and 7B).

```

Client 9 is opened
RTSPSession identifier: 9. Connection arrived from 150.150.100.110:62488.
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 9. Received RTSP message from 150.150.100.110:62488. Message: 'OPTIONS rtsp://150.1
RTSPSession identifier: 9. Sent to client: 'RTSP/1.0 200 OK
RTSPSession identifier: 9. Received RTSP message from 150.150.100.110:62488. Message: 'DESCRIBE rtsp://1
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 9. Is the 'C:\redMUG\Contents\sample.mp4' MP4 file in cache? 'true'
RTSPSession identifier: 9. Video track name to stream: trackID/2, audio track name to stream: trackID/1
RTSPSession identifier: 9. Sent to client: 'RTSP/1.0 200 OK
RTSPSession identifier: 9. Received RTSP message from 150.150.100.110:62488. Message: 'SETUP rtsp://150.1
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 9. Sent to client: 'RTSP/1.0 200 OK
Session: 247776867;timeout=360
RTSPSession identifier: 9. Received RTSP message from 150.150.100.110:62488. Message: 'SETUP rtsp://150.1
Session: 247776867
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 9. Sent to client: 'RTSP/1.0 200 OK
Session: 247776867;timeout=360
RTSPSession identifier: 9. Received RTSP message from 150.150.100.110:62488. Message: 'PLAY rtsp://150.15
Session: 247776867
User-Agent: LibVLC/1.1.2 (LIVE555 Streaming Media v2010.03.16)
RTSPSession identifier: 9. Called the seek method. Start time in secs: 0.000000.
RTSPSession identifier: 9. Sent to client: 'RTSP/1.0 200 OK
Session: 247776867;timeout=360

```

Figure 7A. Connection still active when the data streaming are transforming (server side)



Figure 7B. Video streaming is displayed in the client (behind the NAT)

When negotiation is started, the host records the connection details: IP, TCP port and the streaming UDP ports. The host will insert the fabricated TCP header (after the UDP header) in the video streaming packet before sending it.

The reason for the success of transforming the streaming data is that the sending host converts each streaming UDP packet into a fabricated TCP packet that bypasses the NAT because it uses the *active channel*. The receiving host in turn restores the fabricated TCP packet into the UDP streaming data at the data link layer before sending it to the upper layer.

V. EVALUATION

A comparison between our proposed method and the existing technologies is presented in Table 1. The proposed method has several advantages over the existing technologies, although the new packet size is 20 bytes larger than the normal RTP/UDP packet, but less compared with the HTTP. This has a little impact on the network performance.

The proposed method can traverse the video streaming over all types of NAT. It can also traverse the firewall that blocks the UDP ports that RTP may use, commonly with home

Internet gateway. Streaming might fail at times even if the gateway has a built-in RTSP NAT.

Reference [4] utilizes the two streaming protocols separately by using a third device (proxy) between the client and server (every side with whole streaming protocol advantages and disadvantages), the proposed method utilizes the benefits of the RTP and HTTP protocols without using any extra device.

Table I. CURRENT AND PROPOSED METHOD COMPARISON

FEATURE	HTTP	RTP/TCP	RTP/UDP	PROPOSED METHOD
Directional	Bidirectional	Bidirectional	Uniary	Uniary
Playback hiccups	Yes	Yes	No	No
Quality Reductions	Yes	Yes	No	No
Oscillating throughput	Yes	Yes	No	No
Slow recovery	Yes	Yes	No	No
ISP firewall	Traverse	Traverse	Blocked	Traverse
NAT traversal	Yes	No	No	Yes
End-to-End Delay	Long	Long	Short	Short
Window buffer and reordering	Yes	Yes	No	No
Streaming method	Downloading or progressive	Streaming	Streaming	Streaming

VI. CONCLUSION

The two main transport layer protocols: TCP and UDP can be used in streaming but with the whole advantages and disadvantages of using that protocol. In this paper, a new method is presented and implemented that can merge some advantages of both protocols. It enables the client and server to use UDP advantages in each side for streaming. Both client and server gains scalability by not having to deal with some TCP processing feature (e.g. Acknowledgement and window buffering ...etc). In the other hand, utilize the benefit of the TCP advantages to traverse the NAT and the firewall. In other words, the stream is not discarded and traverses the NAT and the firewall. The experimental results show that the new method achieves the firewall traversal and Nat traversal even with the most difficult NAT (symmetric NAT).

REFERENCES

- [1] Chu-Hsing, L., et al., *Energy Analysis of Multimedia Video Streaming on Mobile Devices*, in Proceedings of the 3rd International Conference and Workshops on Advances in Information Security and Assurance. 2009, Springer-Verlag: Seoul, Korea.
- [2] Matthew Syme, P.G., *Optimizing Network Performance with Content Switching: Server, Firewall and Cache Load Balancing*. 1st ed. 2004: Prentice Hall. 288.
- [3] Philip, W.F., et al., *Server-efficient high-definition media dissemination*, in Proceedings of the 18th international workshop on Network and operating systems support for digital audio and video. 2009, ACM: Williamsburg, VA, USA.

- [4] H, et al., *Transparent protocol translation for streaming*, in Proceedings of the 15th international conference on Multimedia. 2007, ACM: Augsburg, Germany.
- [5] Sourour, M., B. Adel, and A. Tarek. *Security Implications of Network Address Translation on Intrusion Detection and Prevention Systems*. in *Network and Service Security*, 2009. N2S '09. International Conference on. 2009.
- [6] Sanmin, L., et al. *TCPBridge: A software approach to establish direct communications for NAT hosts*. in Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on. 2008.
- [7] Yeryomin, Y., F. Evers, and J. Seitz. *Solving the firewall and NAT traversal issues for SIP-based VoIP*. in Telecommunications, 2008. ICT 2008. International Conference on. 2008.
- [8] Takabatake, T. *A Scheme of Relay Server Selection Methods for NAT Traversal through Home Gateways*. in Telecommunication Networks and Applications Conference, 2008. ATNAC 2008. Australasian. 2008.
- [9] P.Srisuresh, K.E., *Traditional IP Network Address Translator (Traditional NAT)*. RFC 3022, 2001. IETF.
- [10] Khlifi, H., J.C. Gregoire, and J. Phillips, *VoIP and NAT/firewalls: issues, traversal techniques, and a real-world solution*. Communications Magazine, IEEE, 2006. 44(7): p. 93-99.
- [11] Ivanov, I., *API Hooking Revealed*. 2002.
- [12] Hasegawa, T. and T. Ogishi. *A framework on gigabit rate packet header collection for low-cost Internet monitoring system*. in Communications, 2002. ICC 2002. IEEE International Conference on. 2002.
- [13] Jelger, C.S. and J.M.H. Elmirghani. *Performance of a slotted MAC protocol for WDM metropolitan access ring networks under self-similar traffic*. in Communications, 2002. ICC 2002. IEEE International Conference on. 2002.

AUTHORS PROFILE



Omar A. Ibraheem (PhD) is currently a Post Doctoral Research Fellow in National Advanced IPv6 Centre of Excellence (NAv6) at Universiti Sains Malaysia (USM). Dr. Omar obtained his bachelor, master, and doctorate in computer science from Mosul University in 1998, 2000, and 2006 respectively. He has joined NAv6 since January 2010. Before that, Dr. Omar was a senior lecturer at the computer science department, College of Computer Science and Mathematics of Mosul University in Iraq. His research area include the Network protocols, Routing protocols, Network security and Multimedia communications.



Omer Amer Abouabdallah (PhD) is a senior lecturer and post graduate coordinator of the National Advanced IPv6 Centre of Excellence (NAv6) in Universiti Sains Malaysia. Dr. Omar obtained his Bachelor of Computer Science from Al-Fateh University, Tripoli, Libya in 1993. He obtained his Master of Computer Science and doctorate from Universiti Sains Malaysia in 1999 and 2004 respectively. Dr. Omar is heavily involved in researches carried by NAv6 centre, such as IPv6 over Fiber Object and the Multimedia Conferencing System. The highlights of Dr. Omar's achievements include the winner of Sanggar Sanjung Award 2005 and 2006 by USM on 2007 and the winner of Innovative Product Award for NAT AND FIREWALL TRAVERSAL SOLUTION for MSCv6 as well as the gold medal winner in International Innovation Innovation Industrial Design & Technology Exhibition 2006 (ITEX2006).



Sureswaran Ramadass (PhD) is a Professor and the Director of the National Advanced IPv6 Centre (NAv6) at Universiti Sains Malaysia (USM). He is also the founder of Mlabs Systems Berhad (MLABS), a public listed company on the MESDAQ. Prof Dr Sureswaran obtained his BsEE/CE (Magna Cum Laude) and Masters in Electrical and Computer Engineering from the University of Miami in 1987 and 1990 respectively. He obtained his doctorate from USM in 2000 while serving as a full time faculty in the School of Computer Sciences. His research areas include the Multimedia Conferencing System, Distributed Systems and Network Entities, Real Time Enterprise Network Monitoring, Real Time Enterprise System Security, Satellite and Wireless Networks, IPv6 Research, Development and Consultancy, and Digital Library Systems.

The Integration of GPS Navigator Device with Vehicle Tracking System for Rental Car Firms

Omarah Omar Alharaki
Faculty of Information and
Communication Technology,
International Islamic University
Malaysia,
Kuala Lumpur, Malaysia.

Fahad Saleh Alaieri
Faculty of Information and
Communication Technology,
International Islamic University
Malaysia,
Kuala Lumpur, Malaysia.

Akram M. Zeki
Faculty of Information and
Communication Technology,
International Islamic University
Malaysia,
Kuala Lumpur, Malaysia.

Abstract — the aim of this research is to integrate the GPS tracking system (tracking device and web-based application) with GPS navigator for rental cars, allowing the company to use various applications to monitor and manage the cars. This is enable the firms and customers to communicate with each other via the GPS navigator. The system should be developed by applying new features in GPS tracking application devices in vehicles. This paper also proposes new features that can be applied to the GPS Navigator. It also shows the benefits that the customers and staff will get from this system.

Keywords: GPS tracking system, GPS devices, GPRS, Garmin.

I. INTRODUCTION

The Global Positioning System (GPS) is a satellite-based navigation system made up of a network of 24 satellites to specify a position on the surface of the earth [11] [4]. It also provides highly accurate location with the use of special GPS receivers and their augmentations [18]. In 1973 GPS was intended for USA military applications [7], but in the 1980s, the government made the system available for civilian use. GPS works in any weather conditions, anywhere in the world, 24 hours a day. There are no subscription fees or setup charges to use GPS [11].

From NASA (2009) : The uses of GPS have extended to include both the commercial and scientific worlds. Commercially, GPS is used as a navigation and positioning tool in airplanes, boats, cars, and for almost all outdoor recreational activities such as hiking, fishing, and kayaking [7]. GPS is also playing an increasing role in the tracking of motor vehicles [2].

General Packet Radio Service (GPRS) is an enhancement of GSM networks to support packet switched data services such as email and web browser in addition to existing GSM data services such as Short Message Service (SMS). GPRS operates on the existing GSM network infrastructure that it utilizes available time slots during each frame transmission. Thus, it does not

overload the existing GSM network traffic and can efficiently provide data services. The GPRS can transfer data at the maximum rate of 115.2 kbps. Due to a very large coverage area of GSM networks around the world, GPRS becomes the largest data service network available and always-on; thus, it is most suitable for a real-time tracking management system. [1].

GPS tracking system developed that transmit vehicle's data in real time via cellular or satellite networks to a remote computer or data centre [15][10]. Vehicle tracking system signifies the monitoring and management of vehicle, trucks, etc by using GPS system that can get in real time the current location, situation , history, performance and emissions and control them. [1] [16].

The web based tracking system allows users to securely log in and track their cars in real- time over the Internet. The user sees moving dots on a map in a web browser. It also display all transmitted information to the users along with displaying location of vehicle on a map [15].

The tracking system allows users to locate any car at any time of day. Also, it can replay a past trace of the cars history. It can remotely control the car such as run alarms and locking devices. It enable the users to keep track of the vehicles without the intervention of the driver where, as navigation system helps the driver to reach the destination [12].

Many shipment companies in the world use GPS tracking systems in their trucks. It is very important for the fleets, especially when they have big number of trucks and staffs, to manage this huge number of vehicles and people. In addition, The shipments and conveying the tracking information to customers are perceived to be important customer service components and they are often considered industry norms rather than a potential competitive advantage for shipment service providers [14].

However, the biggest challenge in the cars rental companies are car thieves and delay car return. This will cause, lost a lot of money and cars. However, the rental cars firms tried to find the solution to save their cars from these troubles from their clients and staff.

The vehicle tracking system presented in this paper as a system that is designed to track and manage rental cars that are used by the customer, using a GPS tracking technology with GPS navigator.

The system comprises of vehicle GPS tracking devices (Tracking device and GPS navigator), GPRS and a web-based application. Through this system, the company staff will have the facility to monitor the movement and relevant information of each vehicle. Moreover, to keep in touch with their customers by GPS navigator technology [13].

The paper highlights how to apply new application in GPS navigator that allows the firm staff and customers to communicate between each other via GPS navigator in the rental cars. Moreover, satisfying customer needs and building high confidence between the firm and customers.

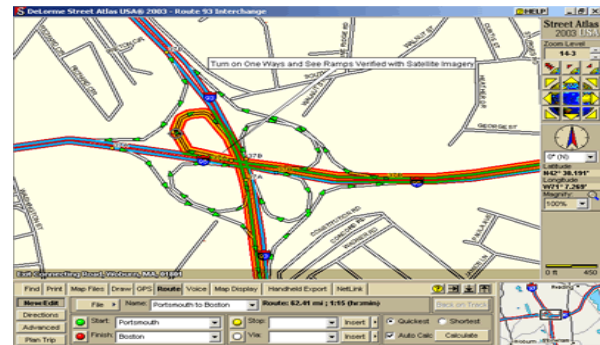
This paper illustrates the integration of multiple technologies to achieve a common goal. It shows how such technologies can be synergistically combined to address real rental cars firm problems [13].

II. GPS TRACKING SYSTEM

GPS device is a small unit that receive signals from satellites and send other signals to antennas (GPRS). This device is a major part of the system and it will be installed into the vehicle which is responsible for capturing the following information for the vehicle such as the Current location of vehicle, Speed of vehicle, Door open/close status, Ignition on/off status, etc. [15].

This device is also responsible for transmitting this information to the Tracking Server located anywhere in the world. Also, it has to install the unit in a hidden and safe place inside the vehicle [15].

The information about the vehicle saved in this unit will be sent to antennas by GPRS, and there are many applications (Figure1 shows one of the application) connected to the internet that can calculate it and put it on the map to integrate with it.



(Figure 1: web based tracking application)

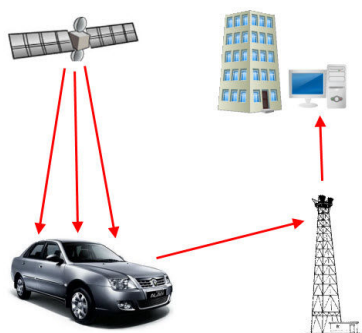
The information includes location, speed, fuel level, engine situation, start driving point, and car situation. Also via this application, the company staff can control the cars, for example, on/off lights, on/off air conditioning, on/off car engine, on/off security system and others, by sending some codes to the tracker in the car.

A. GPS Tracking System Framework

The GPS tracking system consists of client-server architecture where the web browser is the client and the server functions are shared between a web server, a communication server, a database server and a map server [13]. The process of web application tracking consists of four parts: a location-aware web system, location determination, location-dependent content query and personalized presentation. The location-aware web system is the underlying infrastructure. It allows the exchange of the location information between the web browser and the web server, and the automatic update of the web pages in the web browser when the location changes [17].

GPRS is the main method of communication between the tracking device and the web server. GPRS, being a 2.5G mobile technology, is available all over the world. It is also ideally suitable for data transfer over an always on-line connection between a central location and mobile devices. The cost is per kilobyte of data transferred, in comparison to SMS where the cost is per message. [13].

The location information collected through the GPS in real time is placed in a central database that is owned by the firm's staff via GPRS antenna. Each user of the system may access this information via the Internet [15]. Figure 2 shows the system framework.



(Figure 2 : GPS tracking system framework)

B. Features of GPS Tracking System:

Some rental car firms offer GPS devices for rent when a customer comes to rent a car by adding charges for it per day. This is beneficial for customers to locate their rental cars, know the location direction and their current location. Moreover, it also helps the rental car firms to be informed how the customers are using the cars.

This system is also used to prevent theft and retrieve stolen/lost vehicles. The signal sent out by the installed device help the rental car firms to track the vehicle. These tracking systems can be used as an alternative for traditional car alarms or in combination of it. Installing tracking systems can thus bring down the insurance costs for these vehicles by reducing the risk factor [10].

Moreover, this system gives benefits to the rental car firms and to their customers such as: [9].

- Real-time location of vehicles,
- Historical vehicle reports,
- Security code/pin,
- Trip computer,
- Engine idle and start/stop reporting,
- Real-time tracking alerts and reports after hour vehicle monitoring,
- Create custom Geo-fences and landmarks,
- Historical movement of vehicle,
- Mileage reporting,
- Optional starter disable/enable,
- Notification of doors opening and closing

C. Case Study:

Some fleet firms that applied this tracking system which are the following:

Delhi Transport Corporation is the one of the largest City Road Transport Undertaking in India [3]. It has a fleet of around 15,000 vehicles carrying on the business of passenger transport in 800 routes from 33 depots all over the state of Delhi with a product mix comprising of City and Inter-city services. After they implemented the tracking system in vehicles and other functions under Automatic Fleet Management System, the benefits they obtained are [3]:

- Better bus scheduling.
- Quick replacement in case of breakdown/accident en-route.
- Effective control over the drivers & checking bus-stops skipping.
- Check on over-speeding.
- Basic communication between driver and control room in emergencies.
- Automation of Fleet Operations minimizing human intervention.
- Improved Fleet utilization leading to better services and thereby enhancing commuter satisfaction.

Another car rental Firm is in the United Arab Emirates (UAE) that fitted their cars with high-tech C track GPS/GPRS satellite tracking units to prevent thieves from driving away with the cars [8].

Thieves in Dubai and Sharjah are increasingly posing as clients who rent a car and then ship the car out of the country, mostly to Russia, North Africa and Eastern Europe. This prompted car rental firms to integrate vehicle with GPS tracking system to protect their business and save themselves time and money. Al Mumtaz Rent-a-Car in Dubai gives their vehicles some level of safety and it can monitor their vehicle all the time. [8].

The system is cost-effective, as it saves time and money in the recovering of stolen vehicles. According to Diamond Lease, they can also monitor the movement of all their vehicles and can thus establish whether a vehicle is being misused. The GPS tracking device can keep track of the car's engine condition by recording harsh braking, speeding and even the removal of any of the car's parts, thus saving us more money [8].

In Saudi Arabia from rental cars company report they have more than 300 cars and their experience more than 25 years in this field. Moreover, they have more than 10 branches in East of Saudi Arabia. However, during this time the rental cars faced many troubles from either customers or employees such as car lost, personal car use from the staff and delays car return. From these problems the rental cars lose money, lose cars, low service quality, and short cars use period.[one of authors experience]

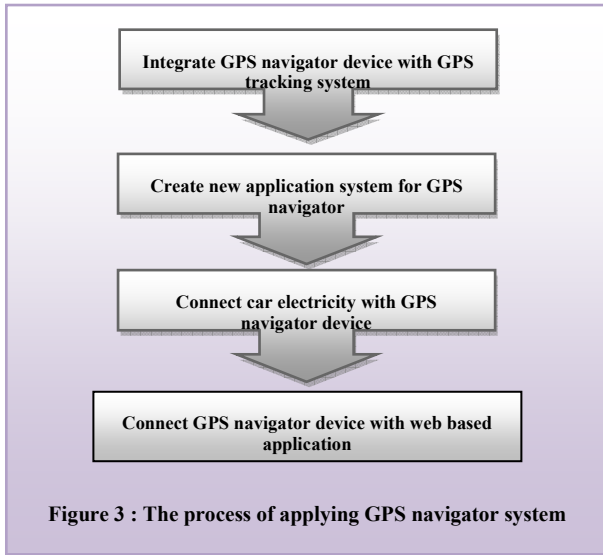
On the other hand, after they installed the tracking system into the rental cars they save almost all of their money, time and cars. Moreover, the firm got improvement of their services quality and maintained the cars quality for a longer period.

III. PROPOSED METHOD AND RECOMMENDATION

This paper shows that this system is easy to apply, very important to manage, save and control the cars. In

this section, the paper will propose some helpful and useful features in the GPS system.

There are some new features recommended to be implement in the system that will save Firm time, money, and manage their cars safely which are as following (Figure 3):



1) *Integrate GPS navigator device with GPS tracking system*

This device usually work as a guide for travelers to show them where they are , where they want to go , maps , roads, shops and other useful information .

The integration framework is between GPS navigator device and GPS tracking system as shown in figure 4.



(Figure 4: framework for the integration of GPS devices)

2) *Create new application system for GPS navigator* that can link customers with rental car company via GPS tracking device. This system will contain new menus that provide new services for customers in addition to its main function as guidance. The proposed new menu can help the customers to contact with staff for any help such as lost directions, lost key or any other assistants as shown in figure 5.



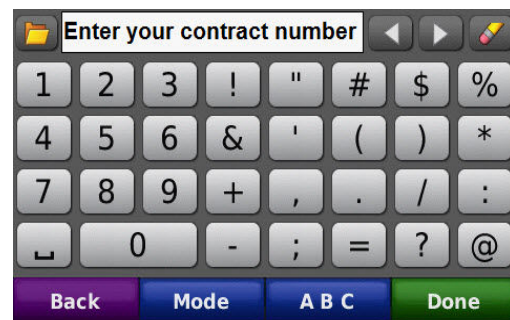
(Figure 5: Proposed new menu in GPS nav. Application)

GPS navigator device must be installed in the cars because it will link with the GPS tracking device.

3) *Connect car electricity with GPS navigator device* that controls the cars, so that nobody can drive the car without logging in using the specific username and password. When the customer wants to drive the car after he/she has signed the contract, he/she has to enter the name that is in the contract and the contract number as shown in figure 6 and figure 7.

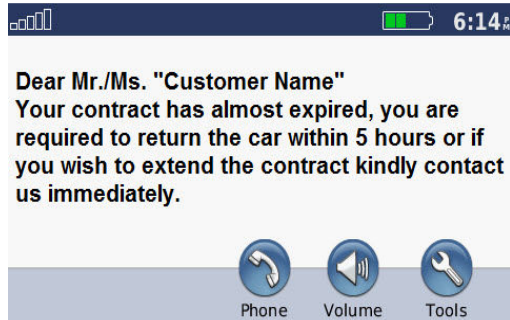


(Figure 6: field for entering customer name)

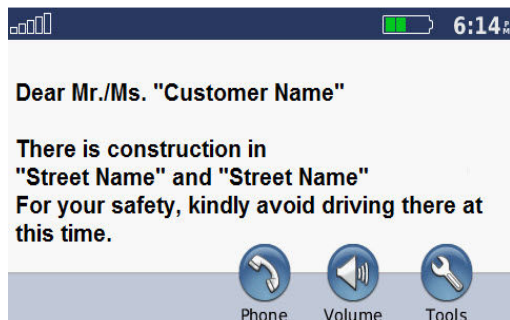


(Figure 7: field for entering contract number)

4) *connect GPS navigator device with web based application* through GPS tracking device that can send messages from the company to the customer, such as, reminders of expiry date of contract, last promotions, and other useful information (as shown in figure 8 & 9) .



(Figure 8: one of the proposed reminder messages from the rental car firm to their customers)



(Figure 9: one of the proposed notice messages from the rental car firm to their customers)

IV. CONCLUSION

In conclusion, this paper presents the development of a Rental cars tracking system using GPS & GPRS technologies. It is a typical example of how the advantages may be forced for the efficient and effective managing of rental cars firm. However, after implementation of this proposed system may give benefits for the rental cars firm customers such as built in road maps, Route capability Touch screen access, monitor fuel consumption, vehicle maintenance alerts, route guidance, Speed limit display [6], last promotions from the firm, warning reminders, renewal car rental contract, and keep contact with the rental cars firm. Moreover, this system show that the customers will connected with the rental cars firm in whole day via GPS navigator device with full protection of the rental cars firms from any violations from the customers.

ACKNOWLEDGMENT

We would like to acknowledge our mentor, Dr. Akram zeki for his patience, guidance, and knowledge in order to reach our goal. Also, we would like to convey our gratitude to our parents who support us to understand the importance of knowledge and show us the best way to achieve it.

REFERENCES

- [1] Chadil, N.; Russameesawang, A.; Keeratiwintakorn, P. (2008), "Real-time tracking management system using GPS, GPRS and Google earth", proceeding of ECTICON.2008, pp.393 – 396.
- [2] Craig A. Scott (1994), "Improved GPS Positioning for Motor Vehicles Through Map, Matching", University of

- Technology, Sydney, Presented at ION-94, Salt Palace Convention Center, Salt Lake City.
- [3] DELHI Transportation Corporation case study (2007), http://www.cmcltd.com/case_studies/transportation/GPS_System_Case_Study_DTC.pdf
- [4] General Information Document of GPS-based Fleet Management and Tracking Systems (2008), Exaterra Inc., Ottawa, Canada: Canadian company based in Ottawa.
- [5] Hariharan, Krumm, Horvitz (2005), "Web-Enhanced GPS", School of Information and Computer Sciences, University of California, Irvine, USA.
- [6] <http://gpstrackit.com/gps-tracking-products/garmin-integration>
- [7] <http://scign.jpl.nasa.gov/learn/gps1.htm>
- [8] <http://www.digicore.com/>, (UAE Press Release August,2007)
- [9] <http://www.imardainc.com/smarttrack-vehicle-tracking-system>
- [10] <http://www.roseindia.net/technology/vehicle-tracking/VehicleTrackingSystems.shtml>
- [11] <http://www8.garmin.com/aboutGPS>
- [12] <http://www.gisdevelopment.net/technology/gps/techgp0044.htm>
- [13] M. Medagama, D. Gamage, L. Wijesinghe, N. Leelaratna, I. Karunaratne and D. Dias (2008), GIS/GPS/GPRS and Web based Framework for Vehicle Fleet Tracking, ENGINEER, No. 05, pp. 28-33.
- [14] Mikko Kaˆrkkˆinen, Timo Ala-Risku, Kary Fraˆmling (2004), "Efficient tracking for short-term multi-company networks", International Journal of Physical Distribution & Logistics Management Vol. 34 No. 7, pp. 545-564
- [15] Muruganandham, P.R.Mukesh (2010), " Real Time Web based Vehicle Tracking using GPS", World Academy of Science, Engineering and Technology 61 2010.
- [16] Robert W. Bogue (2004), "New on-vehicle performance and emission monitoring system", Sensor Review, Volume 24, No.4, pp. 358–360
- [17] Rui Zhou (2008), "Enable web-based tracking and guiding by integrating location-awareness with the world wide web", Campus-Wide Information Systems Vol. 25 No. 5, pp. 311-328
- [18] S.S.S. Prakash, Madhav N. Kulkarni (2003), " Fleet Management: A GPS-GIS integrated approach", Department of Civil Engineering, IIT Bombay, Mumbai, GISdevelopment.net.

Process Framework in Global eXtreme Programming

Ridi Ferdiana, Lukito Edi Nugroho, Paulus Insap
Santoso

Department of Electrical Engineering and Information
Technology
Gadjah Mada University (UGM)
Yogyakarta, Indonesia

Ahmad Ashari

Department of Computer Science and Electronics
Gadjah Mada University (UGM)
Yogyakarta, Indonesia .

Abstract—Software development life cycle works as a process framework that underlying of the software engineering framework. In multi-site software development, the impact of not having a process framework is often quite disastrous. The multi-site team can work anonymous and have no guidance to work and collaborate with others. Therefore, several researches have begun to introduce the important of the process framework through a legitimate software development life cycle. The most common process framework that introduced in multi-site software development is called as a Global Software Development process framework (GSD). However, many GSD implementations are reported as an enterprise process framework that highly iterative and model driven oriented. This paper will show an alternative way to modify the existing GSD SDLC into the simplified process framework by integrating the process framework with an agile method like eXtreme Programming. By simplifying the process framework, it will provide an opportunity for small and medium enterprise to adopt the proposed SDLC in multi-site development.

Keywords—*process framework; software development lifecycle; agile; eXtreme Programming*

I. INTRODUCTION

Global Software development (GSD) is defined as a software development process that uses teams from multiple geographic locations. The physical distant between team is the key issue in GSD. The team may be from the same organization, collaboration that involves different organization, outsourcing to the other organization. The dispersed can happen in same country or other sides of the world. Developing software in the distant introduces complex and interesting issues.

Further research tells us that distributed is done in order to fulfill well-establish motivators include [3].

- Limited trained workforce in technologies that are required to build today's complex systems.
- Differences in development costs favoring dispersing team geographically.
- A "round-clock" work system facilitated by time zone differentials allowing for shorter times to market.

- Advanced infrastructures (internet bandwidth, and collaboration software).
- A desire to be close to a local market.

Although the motivations behind GSD often differ from the other, the main problem is still constant. It is naturally more difficult to coordinate projects where teams separated by physical distance. While in collocated software development there are generally a best practices standard, artifact, and ad-hoc coordination to do "real-time" software development. It does not really exist in distributed software development.

Best practices, artifacts, or ad-hoc communication are done to supply sufficient "shared vision" between the team, and also client. Shared vision is creating a common understanding about what the team is trying to do, what the finished product look like, what the basis of the product, and when they must deliver the product if it is to have its intended effect [8]. Either the system is developed in collocated or distributed; the shared vision its necessity to limit the development risk.

Shared vision in distributed development faces some challenges and issues. The first matter is commonality. It is uncommon to have multiple divisions, organizations, cultures, and languages on a project. Often the team members have not known each other, may have different level experiences, and may have motivation conflict. All these factor plots to make it increasingly difficult to coordinate across teams, manage evolution, and monitor progress. Past studies have shown that tasks take about 2.5 times longer that the collocated one [6].

Sangwan et al. [11] creates a set of practices that can be used on projects that are geographically distributed. The research identified a crucial factor to the success of GSD projects and construct best practices for these factors to enable a successful outcome. The outcome of the research is a process framework that is leveraged the best practices in the critical success factors and is based on the agility model. However, the research still argued about agile implementation as a tradeoff of its discipline process [2].

II. GSD PROCESS FRAMEWORK

Process framework in GSD is divided in four major phases, which are requirements engineering, project planning, architecture design, and product development. Those phases are known as a software development life cycle (SDLC). GSD SDLC provides wide range phases to initiate the project (planning and requirement) and to construct the product (architecture design and product development). Along with the process framework, GSD also provides organization process, and monitoring control as parts in its process framework. This section will discuss the concept in separated sub sections. Figure 1 shows the GSD process framework.

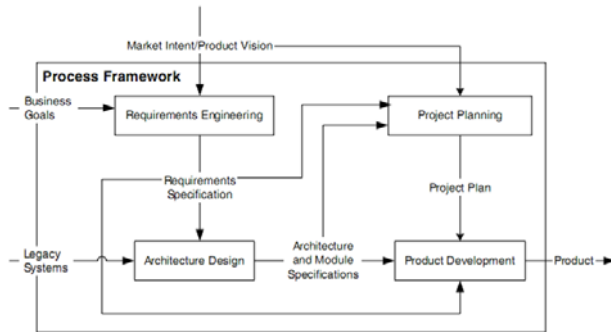


Figure 1. GSD process framework [11]

Each process can be threatened as a process component that has several input and output. Although it shows the sequential one, several implementation of the process framework is described as highly iterative and incremental. The next section will discuss the concept in separated sub sections.

A. GSD requirements engineering

The major concern in requirement engineering is doing analysis of the problem, and creating a system based on the available solution. It can be composed as functional requirements and non-functional requirements. Requirements answer questions about what should be implemented, how the system should behave, what system attributes and constraints.

Requirement engineering process is started by finding business goals and market intent for the product. The product should have a reasonable definition to build, the benefit for the system, and have clear target markets. Without reasonable definition, it can be difficult to define concretely the users, overlooked the product, and create an assumption of the product.

Sangwan, et al. [11] proposes the requirement engineering process through three main activities; which are elicitation, modeling, and review. Each activity provides output and process one or several input.

The elicitation activity is to extract the domain knowledge from the experts, identify features that are in scope and out of scope, and document the features in the highly structured and logical model. In GSD, that activity can be done through several approaches starting from face-to face meeting (it's meaning the team should plan to travel), using CSCW tools like instant messaging, or phone conferences, or indirectly by

using email communication. Those approaches will be discussed further in case studies section.

The primary output of the elicitation activity is a feature-level requirements model. Feature level requirements model focuses in creating a high-level structure of the product feature without detailing and completing the model itself. The creation step to provide the high-level structure by doing function decomposition and categorize the feature through work based structure model. Through this way the GSD team will get start with easy and common understanding through a simple model like Unified Modeling Language.

The model itself provides indirect control between project planning and the design model. However, the requirement model can sometimes become complex, difficult to understand, and have been conflicting with requirements. Therefore, the review process is important to validate the project plan, adjusting the design, and following the software process. Tools are playing some rules to provide several syntactic reviews (e.g. model completeness, violation in rules, etc.). The semantic syntactic should be manually reviewed. In GSD, these step called requirement review meeting. These meeting have three goals, which are; ensuring the requirements were correctly understood and translated into the model, verifying the requirement model itself is readable and understandable, and disseminating the requirements itself.

B. GSD Architecture Designs

The drivers for software architecture in GSD project are shared vision. When the distributed teams have shared vision about the architectures, the development and design will shape by itself. The concerns are many people intuitively know how to create shared vision, but make the intuition explicit about the remote teams is challenging. Shared vision come when the team has same domain knowledge, and is able to design the architecture in well-defined and loosely coupled components.

Define work units is the first step in GSD architecture design to decompose the whole system into functional units that can be allocated distributed in development. The drivers for how to decompose the functional units come from several considerations such as previous experiences, architecturally significant requirements, referenced patterns, organization information, and resource location.

The second step is to identify module responsibility. The functional unit results are combined together with the requirement models. The result of this step is module specification, which is described the detail function of each module. As a structural result, each module will be categorized into two main categories, which are static and dynamic modules. Static modules are defined as a module that works and dedicated for common purposes such as security, print, utility, compression, and any modules that work loosely coupled with the business process. Dynamic modules are defined as a collection of modules that work tight coupled with the business process, e.g. payment processing, inventory checking modules, etc. Both module categories are displayed in a relation through an UML interaction diagrams.

The third steps in architecture development are creating the architecture documents. Architecture document works as a reference guide in development, not as a manual to build the system. Clements et al. [4] provides a guidance to build an architecture document in a sufficient way. According to the research, an architecture document consists of two main views, which are; execution views and implementation views.

C. GSD project planning

Project planning in GSD is focused in allocating the amount work, budget and time in distributed manners. Project planning in GSD usually uses time-boxed development to approach synchronous development among the various geographically distributed development sites. Requirements and architecture results are the main input for this process. Cost estimation and project plan is the main output in project planning.

Feature release planning is the first phases in project planning. It is designed from market intent and functional specification. The feature release planning is an iterative activity that can be revised in time constraints depend on the market intent and requirements changes. Feature release planning is dividing the features of the system through time iteration. Iteration may vary from two weeks or months. The goal of this activity is to identify the minimum set of features that can be sold in the marketplace.

Development planning is proposed by mixing feature release plan, design model, and critical path analysis. The result of the development planning is a project plan. Project plan consists of schedule planning and integration-test planning.

Schedule planning provides a fix date for features. It is driven by the critical path analysis and design model that proposed in architecture sessions. Thus, each development team must release an operational version to integrate and tested on the date specified. If the team can make the date, the release would be skipped, and the team would release on the next fixed date on the next sprint.

Integration and test planning provides a detail when the production can hit the date. It is driven by feature release plan and schedule planning. Iterations will be followed by test and integration planning. The result from this step is a detail works of schedule called as a project plan.

Project plan can be a baseline for the team to make budget proposals. It will become to go or not go to decision. Although it becomes a trade off in the budget versus feature. The project plan can be a live artifact for future reference development. Therefore, GSD proposes three phases in project planning, which are planning in inception phase (requirement phase), planning in elaboration phase (architecture design phase), and planning in construction phase (product development).

D. GSD product development

GSD product development is preliminary started by structuring the team. GSD proposes a hierarchical team structure. It contains a central team, and site team. GSD provides a basic composition of the central team and team site. Basic composition provides the basic structure of the team. There are two communication models between central team

and site team, which are vertical communication and horizontal communication. Horizontal communication is a communication model that exists between team members in one site. Vertical communication happens between two different teams (e.g. central team with the site team) through a proxy communication model. Supplier manager is representative of central team, and R&D manager is representative of site team. Table 1 provides the casual team role in GSD.

TABLE I. GSD TEAM STRUCURES

Role	Responsibility
Product manager	create release planning and decide what feature will be released in a milestone
Program manager	plans the entire lifecycle of a product, managing the requirements, and partitioning component development in team site
Supplier manager	Plans the entire components and manages a remote component development team.
Subject matter expert	Provides expertise on the business that the project is to support. Requirements engineer captures customer requests, develop the analysis model and glossary, and translates analysis model to the requirements document
Requirements engineer	Captures customer requests, develop the analysis model and glossary, and converts analysis model to the requirements document.
Architect	Expresses, designates, and communicates architecture solution, implements an architectural prototype, develops specification, and acceptance tests for product components.
QA Experts	Provisions the automated build/test, test management, and defect tracking systems used by the architecture project team.
R&D Manager	Plans the entire development lifecycle and manage the module development plans.
Designer	Focuses to build user interface and designs modules using appropriate design patterns.
Developer	Gears the module and increment the quality of the codes.

The communication models exist to solve coordination between team and inside the team. Solving a problem in GSD need to know who solve the problem, adjust the schedule, and shift task to the other's team. It works using traditional or the internet based communication tools. In line with engineering and planning activities, preparing an adequate project infrastructure for distributed development is an essential project success factor.

Infrastructure supports in GSD should support accessibility, collaboration, concurrency, processes, awareness, and integration. Accessibility is a characteristic that the tools must be available at all development sites. Collaboration is characteristic that the tools must support both real-time collaboration and asynchronous collaboration. Concurrency refers to the extent to which different parties can collaborate explicitly or implicitly on a single artifact. The process Characteristics focuses in a specific support tool for a software engineering process (e.g. Rational Rose for RUP). Awareness could, for example, be supported by making as much information as possible available in a single location and linking different artifacts for navigability. The last

characteristic which is integration is a nice to have that tools support the integration between different tools through real-time integration or file based integration. Conventional tools may not address all the characteristics; therefore, selecting right tools is the important factor in GSD. Table II provides previous researches in the GSD researches.

TABLE II. GSD TOOLS RELATED RESEARCHES

Authors	Research Topic
Fitzpatrick et al., 2006	Discussing the tools and its implementation for GSD management process through notification and chat.
Pillati et al., 2006	Discussing about software configuration management tools in GSD.
Bass, 2006	Explains monitoring technique for GSD process by using shared mental models approach
Taylor et al., 2006	Discussing the implementation opportunity to adopt Agile in GSD through open collaborative tools.
Thissen et al., 2007	Discussing several alternatives for GSD communication tools such as email, web, and webcasts.

GSD Product development is the last step of GSD process. After this step, product should be delivered to the customer. Product in GSD process is defined as software and its artifacts (e.g. document, user manual, installation guide).

III. GSD PROCESS FRAMEWORK CURRENT ISSUES

The phase in GSD is over when the product delivered and passed the user acceptance test. Although GSD provides a comprehensive process, it does still require software engineering methodology like RUP to provide technical how-to. GSD also needs specific tools to make the process become more effective, some research aware of provide tools like notification chat, web email, open collaborative tools, and software configuration models. Joining GSD with effective software development methodology and sufficient tools absolutely give advantages in the overall software project.

Effective software development is defined as a construction method that focused to deliver the right software products in the right time with the right tools. Developing nowadays software faces challenges that are not those much different from building medieval war ships. There is the "unknown" factor, since many technical problems cannot be understood in their entirety at beginning of the project.

Stober and Hansmann [12] identified those challenges such as complexity of the infrastructure on which IT solutions are built. The challenges arising from international and distributed teams, not yet identify dependencies, exploding costs, and rapidly approaching deadlines, requirements that come up late in the game of decision, unexpected issues to mitigate, and the required knowledge is not sufficient and to make it worse the need for cost-effectiveness is going as a market competition. How do we make software development become effective in GSD, while the team is separated geographically?

Requirements are the key of the effective software development [1]. Through the stable requirements, it is easy to

define distinct milestone, development phases, and acceptances closing phases, although it separates on distant. Unfortunately, requirement changes are usually happened even in the most precisely elaborated software development planning. Therefore, it is important to prepare the team to focus on the ability to adapt quickly to face any kind of forthcoming challenges. Involving the development team as stakeholders, rather than suppliers is a great way to establish an open project structure which responsibility and changes control is shared [12].

Since the key of effective software development is functional software, Agile promises a better way to developing software with a lightweight and adaptive process. Agile provides manifestos and twelve principles that can be seen online at <http://www.agilemanifesto.org/>. Those principles focus in delivering working software through individual and interactions, customer collaboration and responding changes. Agile process implemented through several methods like eXtreme Programming (XP), Scrum, Dynamic Software Development Method (DSDM), Agile modeling, Crystal, Lean, etc. Every method has own unique approaches to build in an agile way.

Relating the agile and GSD is somewhat contradiction Agile needs intensive communication both within the customer and the team, but GSD make a distribution of the team that also contributes to the team dysfunction. Communication between peers becomes important to the team's collective understanding. Remote team members miss these, and consequently, their understanding suffers. Miller [9] shows that when the direct communication doesn't exist in their practices. It will weaken the adopted method. Taylor, et al [13] shows the agile adoption in GSD is just like reinventing the wheel, since many of the Agile GSD experience reports are not giving any additional value in the existing GSD guidance. They also recommend the researcher to create a value or a framework to integrate Agile in GSD context.

Although it has a contradiction in the term of conditions, several researchers also provide field reports about successful agile adoption in GSD. Miller [9] confidently said that his team at Microsoft delivered sufficient software by integrating Scrum, XP, and GSD. Hazzan and Dubinsky [5] states the diversity that exists in GSD is naturally supported by agile software development. Paasivaara et al [10] captures the Scrum practices that successfully adopted in three GSD projects. The following section will discuss how eXtreme Programming method is integrated with the current GSD framework to provide a process framework that the research called as the Global eXtreme Programming process framework.

IV. RESEARCH METHODOLOGY

In order to integrate between the GSD process and XP methodology, the research did case studies that based on the real project. The organization that we are selected is the small independent software vendor that contains five development members with two clients who dedicated to support the project. The project length is six months and works to develop project management that running on the web.

The team itself has two years average of experience in multi-site development. The client and the team are separated in the different countries with different time zone. The team composition as like follows.

TABLE III. TEAM STRUCTURES

Roles	Job Descriptions
Customer	Drive the product direction to the team, answer the developer questions, write stories, declare stories complete, write acceptance test, accept the release
Coach	A thorough understanding of the XP process, professional development experience, leadership experience, get the developers and testers to adhere to the values and practices, assume a leadership role for the development and testing teams, pair with other developers and testers as needed, assist customers in the story writing process, assist customers in the acceptance test writing process
Developer	Estimate stories, brainstorm tasks, develop unit tests, develop the tasks, refactor the code, and communicate with customers when questions arise. There are two developers in this case study
Client Subject matter expert	Help the customers define intelligent stories, act as a proxy for the customer.
Tester	Ensure that a story is testable, assist the customers with writing acceptance tests, run the acceptance tests
The Tracker	Collect development metrics, produce reports indicating the team's progress, communicate the team's historical velocity, and communicate the status of the team's progress

The research will collect the data by exploring the case study related artifacts in the proposed workflow. The research procedures are come as follows.

- The team is incubating and does workshops to implement the proposed GXP framework process.
- Based on the theory, the team designs the artifacts.
- The evaluator evaluates the execution and captures several data that can be written as finding of the case study research.

Based on the project execution the GXP process framework finding is discussed the next sections.

V. GXP PROCESS FRAMEWORK

Understanding the GXP process framework understands the hybrid approaches that integrated between conventional GSD process and XP method. As we know, conventional GSD process divides the SDLC into four phases, which are requirements engineering, project planning, architecture design, and production. XP method adopts exploration, planning, iteration, production, and maintenance as a phase in a development cycle. The GXP SDLC proposes the uses the XP method as a baseline and integrates the GSD workflow on it. Figure 2 shows the GXP process framework.

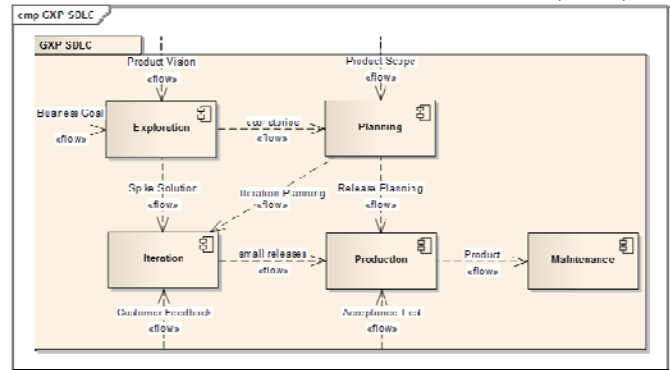


Figure 2. GXP Process Framework

GXP process framework provides several input and output such as.

- **Product vision.** Product vision works as an input that shows a reasonable reason definition why the company building the product, what system will be used for, and what the target markets are.
- **Business goal.** Business goal works as an input that described a business background for the product. It can be related with productivity enhancement, increasing revenue, gaining more customers, or cutting costs.
- **Product scope.** Product scope works as an input that show what kind of feature that will be developed in a timeframe. This input usually depends on available resources, existing budget, and time limitation.
- **User stories.** User stories work as an output from the exploration phases and an input for planning phase. User stories describe the feature level requirements model. User story details are described in user story estimation. The detail's version of user stories is called as the detail requirements model.
- **Spike solution.** Spike solution is a high level architecture of the product. It works as an input for iteration phase and an output from exploration phase. Spike solution also describes the simple design for the product.
- **Release planning.** Release planning is a milestone plan for the product. It described the feature that will be delivered in iteration. Release planning is an output from the planning phases and works as an input for production phase.
- **Iteration planning.** Iteration planning is a set of activities to build features in iteration. The activity is designed from the task card that structured based on planning game activity. It works as an output from planning phase and an input for iteration phase.
- **Customer feedback.** Customer feedback is a customer feedback based on iteration. It described validation and verification for a customer request. Changes will happen when the team built misleading feature. It works as an input for iteration phase.

- Small releases are an output from iteration. Small releases described potential working feature that delivers in iteration. Small releases also work as an input for production phase.
- Acceptance test is an input for production phase. It describes a black box test for product features. The black box test usually runs by the customer.
- Product is an output from production phase and an input from maintenance phase. It works as a final release of the feature.

A. Exploration phase

Exploration phase is where discovery of vision and mission occurs. This phase is equal with requirements engineering process in the GSD. At this stage, the stakeholder with coming up with the vision statement. Vision statement is a high-level statement that describes the goal of the system. For example, the purpose of this new system is to read books via the Internet. Vision statement is not more than 20-30 words. Then, this vision statement is described visually through unique XP artifacts called system metaphor. System metaphor is how the team conceptualizes the system and is typically written in language that is relevant to the business domain.

The detail of the system metaphor is described through a user story. It is the tool that captures user requirements. User stories are similar to use cases and are written by the customer in non-technical language. Most user stories are around 1-4 sentences long. For example, library member can borrow 10 books for a maximum. User stories are written in index card or sticky note that posted on a whiteboard.

Since the GXP is working in multi-site development the user stories and vision statement should be documented in an artifact. GXP proposes the user stories' artifacts such as follows.

- Users, persona or actor that will use the system.
- User stories, the lists of the user scenarios that supported by the system.
- Task cards, it provides a detail tasks that should be performed for each story.
- Software estimation. The software estimation can be done through estimation techniques such as user stories points, or early estimation by the developer.

The user stories' artifacts are composed by the project manager and the clients. The artifacts will be uploaded and maintained through the document repositories that run on the internet.

B. Planning Phase

Planning phase starts with the activity called planning game. Planning game is a short workshop that plays by customer and development team. Customer purposes are determining the value, and development team determines the cost. There are two stages of planning game, the first one is releasing planning game, and the second is iteration planning

game. In the release planning game, the goal is to define the set of features in the next release. The customer writes the stories, development team estimates the story point, and customer plans the overall release. In iteration planning game, the customer chooses the stories for the iteration, and the development team estimate how long, and accepts the corresponding task. Estimation is based on experiences or intuition of the development team. The goal in iteration planning game is to define how long the iteration and what kind of feature that delivered in iteration.

The planning game execution actually runs parallel with the exploration phase. Therefore, the information in user story artifact is refined in the planning phase. This phase also discusses about creates an artifact called Iteration and Release Planning. This artifact is created with a composition as follows.

- Categorized user story. In this section team should create user stories that already categorized based on its urgency. GXP introduces four levels of priority, which are high (essentially needed), medium (adding business value), low (nice to have) and none (verbose). Team can easily throw none level and negotiate for low level.
- Project plan summary. It contains how long is the project, how long the iteration length, how much the user stories, and how many iterations in a project.
- Iteration detail. It describes the user story that executed in the iteration. Actually, the iteration detail discusses the work assignment for developer.

The artifact is also stored and maintained by the project manager. It will be updated by the developer when the user story is completed

C. Iteration Phase

Iteration phase is the real work of development happens. The development team selects the story in current iteration. Since user stories are generally too high-level to execute as programming tasks, so the next step is to break out the actual development tasks needed to support the story. Development team works together to investigate the tasks for each user story. The tasks are written either on the back of the story, on separate task cards, or on some other tracking mechanism. Developers begin work on a new task by writing the test first and adding it to their test framework. Programming continues in pairs with each partner taking turns to "drive" the keyboard from time to time. At the end of the iteration, customers perform the acceptance tests they have written. Any user stories that fail acceptance will be fixed during the next iteration; there is why the iteration phase provides feedback to planning phase.

The result of the iteration is software that works in staging environment. Staging environment is the customer's software environment that mirrors, or replicates, their live environment. Staging is used to perform final prerelease tests and performance checks. Customer will evaluate the staging system and then approve the system shift to the production environment.

In the iteration phase, the artifacts that live in the internet are stored in the source code control systems. The tools like

Team Foundation Server, Team City, or hosted source code online in third party are the several options to store the most important artifacts in the software development. In GXP the developer should have several agreements based on the codes such as follows.

- The coding standard and naming conventions of the codes. This little agreement gives benefits for the other's team to understand the code better.
- The comment agreement. This approach clearly makes the codes speak what developer did with the codes. Having automatic commenting software like GhostDoc, will make it happen.
- The developer notes. This approach makes an agreement how the developer communicate each other through the special comment token. A word like TODO, BUGBUG or DEBUG will make the developer communication is better in term exchanging information.

D. Production and Maintenance Phase

Production is the user acceptance test that happens with the developed codes or modules. The developer set up the system in the development system and tests it by the client. The approved features is copied into the production system or staging system. The approval model works as follows.

- Developers build the codes based on the user stories, and then uploaded into the development system.
- The client and developer do online meeting and discuss through shared screen and review sheets.
- The client fills the review sheet approves it or rejects it for further modification.

After the production, the system goes into the maintenance mode. In this mode, development team can create an upgrade, patch, or changes code with confidence since the change's monitoring can be easily defined through test case. The hard part in this phase is creating data migration and system migration. Since those parts are technology specific, XP does not provide the sufficient techniques to overcome it. The artifact that exists in this phase is.

- Review sheet for user acceptances tests
- Defects document that fulfilled also by the client

VI. DISCUSSION AND FUTURE WORKS

In our studies, we do a novel approaches to integrate the eXtreme Programming with the Global Software Development process. Although both are contradict in the first vision, both can be integrated through the power of artifacts and the tools. The gap in direct communication between client and customer can be limited by using a formal process framework and the tools like source code version document management, or others. The research also notes several finding that related with the artifacts that designed by the team to tackle a communication gaps between of them.

In the future, evaluation can be further improvements of the research. For example, it will be a good idea to compare the proposed process framework with the existing one. The impact of the process framework in terms of productivity, product quality, and resources will make the benefit of the framework can be easily quantified. We are currently in the process of defining some empirical studies with real multi-site distributed projects.

REFERENCES

- [1] Belford, P. C., Bond, A. F., Henderson, D. G., and Sellers, L. S. 1976. Specifications a key to effective software development. In Proceedings of the 2nd international Conference on Software Engineering (San Francisco, California, United States, October 13 - 15, 1976). International Conference on Software Engineering. IEEE Computer Society Press, Los Alamitos, CA, 71-79.
- [2] Boehm, B. and Turner, R. 2004. Balancing Agility and Discipline: Evaluating and Integrating Agile and Plan-Driven Methods. In Proceedings of the 26th international Conference on Software Engineering (May 23 - 28, 2004). International Conference on Software Engineering. IEEE Computer Society, Washington, DC, 718-719.
- [3] Carmel, E. 1999 Global Software Teams: Collaborating Across Borders and Time Zones. Prentice Hall PTR.
- [4] Clements, P., Bachmann, F., Bass, L., Garlan, D., Ivers, J., Little, R., Nord, R., and Stafford, J. 2002. Documenting Software Architectures: Views and Beyond, Addison-Wesley.
- [5] Hazzan, O. and Dubinsky, Y. 2008. Agile Software Engineering. Springer.
- [6] Herbsleb, J. D. and Mockus, A. 2003. An Empirical Study of Speed and Communication in Globally Distributed Software Development. IEEE Trans. Softw. Eng. 29, 6 (Jun. 2003), 481-494.
- [7] Herbsleb, J. D., Paulish, D. J., and Bass, M. 2005. Global software development at siemens: experience from nine projects. In Proceedings of the 27th international Conference on Software Engineering (St. Louis, MO, USA, May 15 - 21, 2005). ICSE '05. ACM, New York, NY, 524-533.
- [8] McCarthy, J. and McCarthy, M. 2006 Dynamics of Software Development. Second. Microsoft Press.
- [9] Miller, A. 2008. Distributed Agile Development at Microsoft Patterns and Practices. Microsoft.
- [10] Paasivaara, M., Durasiewicz, S., and Lassenius, C. 2009. Using Scrum in Distributed Agile Development: A Multiple Case Study. In Proceedings of the 2009 Fourth IEEE international Conference on Global Software Engineering (July 13 - 16, 2009). ICGSE. IEEE Computer Society, Washington, DC, 195-204.
- [11] Sangwan, R., Bass, M., Mullick, N., Paulish, D. J., and Kazmeier, J. 2007. Global Software Development Handbook (Auerbach Series on Applied Software Engineering Series). Auerbach Publications.
- [12] Stober, W. and Hansmann, U. 2010. Agile Software Development: Best Practices for Large Software Development Projects. Springer.
- [13] Taylor, P. S., Greer, D., Sage, P., Coleman, G., McDaid, K., and Keenan, F. 2006. Do agile GSD experience reports help the practitioner?. In Proceedings of the 2006 international Workshop on Global Software Development For the Practitioner (Shanghai, China, May 23 - 23, 2006). GSD '06. ACM, New York, NY, 87-93.

AUTHORS PROFILE

Ridi Ferdiana. Mr. Ridi Ferdiana was born in 1983. He is a doctoral student at Gadjah Mada University, Yogyakarta since 2008. He earned his master degree from the same university in 2006. In his professional area, he holds several professional certifications such as MCP, MCTS, MCPD, MCITP, MVP and MCT. In his daily research activities he really enjoys to learn about software engineering, business platform collaboration, and programming optimization.

Lukito Edi Nugroho. Born in 1966, Dr. Lukito Edi Nugroho is an Associate Professor in the Department of Electrical Engineering and Information Technology, Gadjah Mada University. He obtained his M.Sc. and PhD degrees from James Cook University in 1995 and Monash University in 2002, respectively. His areas of interest include software engineering, distributed and mobile computing, and application of ICT in education.

Paulus Insap Santosa. Insap was born in Klaten, 8 January 1961. He obtained his undergraduate degree from Universitas Gadjah Mada in 1984, master degree from University of Colorado at Boulder in 1991, and doctorate degree from National University of Singapore in 2006. His research interest including Human Computer Interaction and Technology in Education.

Ahmad Ashari Place and date of birth: Surabaya, May 2nd 1963. Get Bachelor's degree 1988 in Electronics and Instrumentation, Physics department Gadjah Mada University, Yogyakarta. Master degree 1992 in Computer Science, University of Indonesia, Jakarta Doctor Degrees 2001 in Informatics, Vienna University of Technology. Major Field of study is distributed system, Internet, Web Services, and Semantic Web.

A Hybrid PSO–SVM Approach for Haplotype Tagging SNP Selection Problem

Min-Hui Lin

Department of Computer Science and Information
Engineering, Dahan Institute of Technology,
Sincheng, Hualien County, Taiwan, Republic of China

Chun-Liang Leu

Department of Information Technology, Ching Kuo
Institute of Management and Health,
Keelung, Taiwan, Republic of China

Abstract—Due to the large number of single nucleotide polymorphisms (SNPs), it is essential to use only a subset of all SNPs called haplotype tagging SNPs (htSNPs) for finding the relationship between complex diseases and SNPs in biomedical research. In this paper, a PSO-SVM model that hybridizes the particle swarm optimization (PSO) and support vector machine (SVM) with feature selection and parameter optimization is proposed to appropriately select the htSNPs. Several public datasets of different sizes are considered to compare the proposed approach with other previously published methods. The computational results validate the effectiveness and performance of the proposed approach and the high prediction accuracy with the fewer htSNPs can be obtained.

Keywords : Single Nucleotide Polymorphisms (SNPs), Haplotype Tagging SNPs (htSNPs), Particle Swarm Optimization (PSO), Support Vector Machine (SVM).

I. INTRODUCTION

The large number of single nucleotide polymorphisms (SNPs) in the human genome provides the essential tools for finding the association between sequence variation and complex diseases. A description of the SNPs in each chromosome is called a haplotype. The string element of each haplotype is 0 or 1, where 0 denotes the major allele and 1 denotes the minor allele. The genotype is the combined information of two haplotypes on the homologous chromosomes and is prohibitively expensive to directly determine the haplotypes of an individual. Usually, the string element of a genotype is 0, 1, or 2, where 0 represents the major allele in homozygous site, 1 represents the minor allele in homozygous site, and 2 is in the heterozygous site. The genotyping cost is affected by the number of SNPs typed. In order to reduce this cost, a small number of haplotype tagging SNPs (htSNPs) which predicts the rest of SNPs are needed.

The haplotype tagging SNP selection problem has become a very active research topic and is promising in disease association studies. Several computational algorithms have been proposed in the past few years, which can be divided into

two categories: block-based and block-free methods. The block-based methods [1-2] firstly partition human genome into haplotype blocks. The haplotype diversity is limited and then subsets of tagging SNPs are searched within each haplotype block. A main drawback of block-based methods is that the definition of blocks is not a standard form and there is no consensus about how these blocks should be partitioned. The algorithmic framework for selecting a minimum informative set of SNPs avoiding any reference to haplotype blocks is called block-free methods [3]. In the literature [4-5], feature selection technique was adopted to solve for the tagging SNPs selection problem and achieved some promising results.

Feature selection algorithms may be widely categorized into two groups: the filter approach and the wrapper approach. The filter approach selects highly ranked features based on a statistical score as a preprocessing step. They are relatively computationally cheap since they do not involve the induction algorithm. Wrapper approach, on the contrary, directly uses the induction algorithm to evaluate the feature subsets. It generally outperforms filter method in terms of classification accuracy, but computationally more intensive. Support Vector Machine (SVM) [6] is a useful technique for data classification. A practical difficulty of using SVM is the selection of parameters such as the penalty parameter C of the error term and the kernel parameter γ in RBF kernel function. The appropriate choice of parameters is to get the better generalization performance.

In this paper, a hybrid PSO-SVM model that incorporates the Particle Swarm Optimization (PSO) and Support Vector Machine (SVM) with feature selection and parameter optimization is proposed to appropriately select the htSNPs. Several public benchmark datasets are considered to compare the proposed approach with other published methods. Experimental results validate the effectiveness of the proposed approach and the high prediction accuracy with the fewer htSNPs can be obtained. The remainder of the paper is organized as follows: Section 2 introduces the problem formulation. Section 3 describes the PSO and SVM classifier. In Section 4, the particle representation, fitness measurement, and the proposed hybrid system procedure are presented. Three public benchmark problems are used to validate the proposed

approach and the comparison results are described in Section 5. Finally, conclusions are made in Section 6.

II. PROBLEM FORMULATION

As shown in Figure 1, assume that dataset U consists of n haplotypes $\{h_i\}_{1 \leq i \leq n}$, each with p different SNPs $\{S_j\}_{1 \leq j \leq p}$, U is $n \times p$ matrix. Each row in U indicates the haplotype h_i and each column in U represents the SNP S_j . The element $d_{i,j}$ denotes the j -th SNP of i -th haplotype, $d_{i,j} \in \{0,1\}$. Our goal is to determine a minimum size g set of selected SNPs (htSNPs) $V = \{v_k\}$, $k \in \{1, 2, \dots, p\}$, $g = |V|$, in which each random variable v_k corresponding to the k -th SNP of haplotypes in U , to predict the remaining unselected ones with a minimum prediction error. The size of V is smaller than a user-defined value R ($g \leq R$), and the selected SNPs are called haplotype tagging SNPs (htSNPs) while the remaining unselected ones are named as tagged SNPs. Thus, the selection set V of htSNPs is based on how well to predict the remaining set of the unselected SNPs and the number g of selected SNPs is usually minimized according to the prediction error by calculating the leave-one-out cross-validation (LOOCV) experiments [7].

	S_1	S_2	\dots	S_j	\dots	S_{p-1}	S_p
h_1	$d_{1,1}$	$d_{1,2}$	\dots	$d_{1,j}$	\dots	$d_{1,p-1}$	$d_{1,p}$
h_2	$d_{2,1}$	$d_{2,2}$	\dots	$d_{2,j}$	\dots	$d_{2,p-1}$	$d_{2,p}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
h_i	$d_{i,1}$	$d_{i,2}$	\dots	$d_{i,j}$	\dots	$d_{i,p-1}$	$d_{i,p}$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots	\vdots
h_{n-1}	$d_{n-1,1}$	$d_{n-1,2}$	\dots	$d_{n-1,j}$	\dots	$d_{n-1,p-1}$	$d_{n-1,p}$
h_n	$d_{n,1}$	$d_{n,2}$	\dots	$d_{n,j}$	\dots	$d_{n,p-1}$	$d_{n,p}$

$n \times p$

Figure 1 The haplotype tagging SNP Selection Problem.

III. RELATED WORKS

A. Particle Swarm Optimization

The PSO is a novel optimization method originally developed by Kennedy and Eberhart [8]. It models the processes of the sociological behavior associated with bird flocking and is one of the evolutionary computation techniques. In the PSO, each solution is a 'bird' in the flock and is referred to as a 'particle'. A particle is analogous to a chromosome in GA. Each particle traverses the search space looking for the global optimum. The basic PSO algorithm is as follow:

$$v_{id}^{k+1} = w \cdot v_{id}^k + c_1 \cdot r_1 \cdot (p_{id}^k - x_{id}^k) + c_2 \cdot r_2 \cdot (g_{best}^k - x_{id}^k) \quad (1)$$

$$x_{id}^{k+1} = v_{id}^{k+1} + x_{id}^k \quad (2)$$

where $d = 1, 2, \dots, D$, $i = 1, 2, \dots, S$, and D is the dimension of the problem space, S is the size of population, k is the iterative times; v_{id}^k is the i -th particle velocity, x_{id}^k is the current particle solution, p_{id}^k is the i -th particle best (p_{best}) solution achieved so far; g_{best}^k is the global best (g_{best}) solution obtained so far by any particle in the population; r_1 and r_2 are random values in the range $[0,1]$, both of c_1 and c_2 are learning factors, usually $c_1 = c_2 = 2$, w is a inertia factor. A large inertia weight facilitates global exploration, while a small one tends to local exploration. In order to achieve more refined solution, a general rule of thumb suggests that the initial inertia value had better be set to the maximum $w_{max} = 0.9$, and gradually down to the minimum $w_{min} = 0.4$.

According to the searching behavior of PSO, the gbest value will be an important clue in leading particles to the global optimal solution. It is unavoidable for the solution to fall into the local minimum while particles try to find better solutions. In order to allow the solution exploration in the area to produce more potential solutions, a mutation-like disturbance operation is inserted between Eq. (1) and Eq. (2). The disturbance operation random selects k dimensions ($1 \leq k \leq$ problem dimensions) of m particles ($1 \leq m \leq$ particle numbers) to put Gaussian noise into their moving vectors (velocities). The disturbance operation will affect particles moving toward to unexpected direction in selected dimensions but not previous experience. It will lead particle jump out from local search and further can explore more un-searched area.

According to the velocity and position updated formula mentioned above, the basic process of the PSO algorithm is given as follows:

- 1.) Initialize the swarm by randomly generating initial particles.
- 2.) Evaluate the fitness of each particle in the population.
- 3.) Compare the particle's fitness value to identify the both of p_{best} and g_{best} values.
- 4.) Update the velocity of all particles using Equation (1).
- 5.) Add disturbance operator to moving vector (velocity).
- 6.) Update the position of all particles using Equation (2).
- 7.) Repeat the Step 2 to Step 6 until a termination criterion is satisfied (e.g., the number of iteration reaches the pre-defined maximum number or a sufficiently good fitness value is obtained).

The authors [8] proposed a discrete binary version to allow the PSO algorithm to operate in discrete problem spaces. In the binary PSO (BPSO), the particle's personal best and global best is updated as in continuous value. The major different between discrete PSO with continuous version is that velocities of the particles are rather defined in terms of probabilities that a bit whether change to one. By this definition, a velocity must be restricted within the range $[V_{min}, V_{max}]$. If $v_{id}^{k+1} \notin (V_{min}, V_{max})$ then $v_{id}^{k+1} = \max(\min(V_{max}, v_{id}^{k+1}), V_{min})$. The new particle position is calculated using the following rule:

$$\text{If } \text{rand}() < S(v_{id}^{k+1}), \text{ then } x_{id}^{k+1} = 1; \text{ else } x_{id}^{k+1} = 0 \quad (3)$$

$$\text{, where } S(v_{id}^{k+1}) = \frac{1}{1 + e^{-v_{id}^{k+1}}} \quad (4)$$

The function $S(v_{id})$ is a sigmoid limiting transformation and $\text{rand}()$ is a random number selected from a uniform distribution in $[0, 1]$. Note that the BPSO is susceptible to sigmoid function saturation which occurs when velocity values are either too large or too small. For a velocity of zero, it is a probability of 50% for the bit to flip.

B. Support Vector Machine Classifier

SVM starts from a linear classifier and searches the optimal hyper-plane with maximal margin. The main motivating criterion is to separate the various classes in the training set with a surface that maximizes the margin between them. It is an approximate implementation of the structural risk minimization induction principle that aims to minimize a bound on the generalization error of a model.

Given a training set of instance-label pairs $(x_i, y_i), i = 1, 2, \dots, m$ where $x_i \in R^n$ and $y_i \in \{+1, -1\}$. The generalized linear SVM finds an optimal separating hyper-plane $f(x) = \langle w \cdot x \rangle + b$ by solving the following optimization problem:

$$\text{Minimize}_{w, b, \xi} \quad \frac{1}{2} w^T w + C \sum_{i=1}^m \xi_i \quad (5)$$

$$\text{Subject to: } y_i (\langle w \cdot x_i \rangle + b) + \xi_i - 1 \geq 0, \xi_i \geq 0$$

where C is a penalty parameter on the training error, and ξ_i is the non-negative slack variables. This optimization model can be solved using the Lagrangian method, which maximizes the same dual variables Lagrangian $L_D(\alpha)$ (6) as in the separable case.

$$\text{Maximize}_{\alpha} \quad L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle x_i \cdot x_j \rangle \quad (6)$$

$$\text{Subject to: } 0 \leq \alpha_i \leq C, i = 1, 2, \dots, m \text{ and } \sum_{i=1}^m \alpha_i y_i = 0$$

To solve the optimal hyper-plane, a dual Lagrangian $L_D(\alpha)$ must be maximized with respect to non-negative α_i under the constraint $\sum_{i=1}^m \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$. The penalty parameter C is a constant to be chosen by the user. A larger value of C corresponds to assigning a higher penalty to the errors. After the optimal solution α_i^* is obtained, the optimal hyper-plane parameters w^* and b^* can be determined. The optimal decision hyper-plane $f(x, \alpha^*, b^*)$ can be written as:

$$f(x, \alpha^*, b^*) = \sum_{i=1}^m y_i \alpha_i^* \langle x_i \cdot x \rangle + b^* = \langle w^* \cdot x \rangle + b^* \quad (7)$$

Linear SVM can be generalized to non-linear SVM via a mapping function Φ , which is also called the kernel function, and the training data can be linearly separated by applying the linear SVM formulation. The inner product $(\Phi(x_i) \cdot \Phi(x_j))$ is calculated by the kernel function $k(x_i, x_j)$ for given training data. By introducing the kernel function, the non-linear SVM (optimal hyper-plane) has the following forms:

$$\begin{aligned} f(x, \alpha^*, b^*) &= \sum_{i=1}^m y_i \alpha_i^* \langle \Phi(x) \cdot \Phi(x_i) \rangle + b^* \\ &= \sum_{i=1}^m y_i \alpha_i^* k(x, x_i) + b^* \end{aligned} \quad (8)$$

Though new kernel functions are being proposed by researchers, there are four basic kernels as follows.

$$\bullet \text{ Linear: } k(x_i, x_j) = x_i^T x_j \quad (9)$$

$$\bullet \text{ Polynomial: } k(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (10)$$

$$\bullet \text{ RBF: } k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (11)$$

$$\bullet \text{ Sigmoid: } k(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (12)$$

where γ, r and d are kernel parameters. Radial basis function (RBF) is a common kernel function as Eq. (11). In order to improve classification accuracy, the kernel parameter γ in the kernel function should be properly set.

IV. METHODS

As the htSNPs selection problem mentioned above in Section 2, the notations and definitions are used to present our proposed method. In the dataset U of $n \times p$ matrix, each row (haplotypes) can be viewed as a learning instance belonging to a class and each column (SNPs) are attributes or features based on which sequences can be classified into class. Given the values of g htSNPs of an unknown individual x and the known full training samples from U , a SNP prediction process can be treated as the problem of selecting tagging SNPs as a feature selection problem to predict the non-selected tagging SNPs in x . Thus, the tagging SNPs selection can be transformed to solve for a binary classification of vectors with g coordinates by using the support vector machine classifier. Here, an effective PSO-SVM model that hybridizes the particle swarm optimization and support vector machine with feature selection and parameter optimization is proposed to appropriately select the htSNPs. The particle representation, fitness definition, disturbance strategy for PSO operation and system procedure for the proposed hybrid model are described as follows.

A. Particle Representation

The RBF kernel function is used in the SVM classifier to implement our proposed method. The RBF kernel function requires that only two parameters, C and γ should be set. Using the RBF kernel for SVM, the parameters C , γ and SNPs viewed as input features which must be optimized

simultaneously for our proposed PSO-SVM hybrid system. The particle representation consists of three parts including: C and γ are the continuous variables, and the SNPs mask are the discrete variables. For the feature selection, if n_f features are required to decide which features are chosen, then $n_f + 2$ decision variables in each particle must be adopted.

Table 1 shows the particle representation of our design. The representation of particle i with dimension of $n_f + 2$, where n_f is the number of SNPs (features) that varies from different datasets. $x_{i,1} \sim x_{i,n_f} \in \{0,1\}$ denotes the SNPs mask, x_{i,n_f+1} indicates the parameter value C , x_{i,n_f+2} represents the parameter value γ . If $x_{i,k} = 1, k = 1, 2, \dots, n_f$ represents the k -th SNP on the i -th particle to be selected, and vice versa.

TABLE I. The particle i representation.

Discrete-variables	Continuous-variables	
SNPs mask	C	γ
$x_{i,1} \ x_{i,2} \ \dots \ x_{i,n_f}$	x_{i,n_f+1}	x_{i,n_f+2}

A random key encoding method [9] is applied in the PSO algorithm. Generally, random key encoding is used for an order-based encoding scheme where the value of random key is the genotype and the decoding value is the phenotype. Note that the particle in each $\{x_{i,k} \}_{1 \leq k \leq n_f}$ is assigned a random number on $(0, 1)$, and to decode in ascending order with regard to its value. In the PSO learning process, the particle to be counted larger tends to evolve closer to 1 and those to be counted smaller tends to evolve closer to 0. Therefore, a repair mechanism such as particle amendment in [5] to guarantee the number of htSNPs after update process in PSO is not required.

B. Fitness Measurement

In order to compare the performance of our proposed approach with other published methods SVM/STSA in [4] and BPSO in [5], the leave-one-out cross validation is used to evaluate the quality of fitness measurement. The prediction accuracy is measured as the percentage of correctly predicted SNP values on non-selected SNPs. In the LOOCV experiments, each haplotype sequence is removed one by one from dataset U , the htSNPs are selected using only the remaining haplotypes to predict these tagged SNPs values for the removed one. This procedure is repeated such that each haplotype in U is run once in turn as the validation data.

C. The Proposed Hybrid System Procedure

Figure 2 shows the system architecture of our proposed hybrid model. Based on the particle representation and fitness

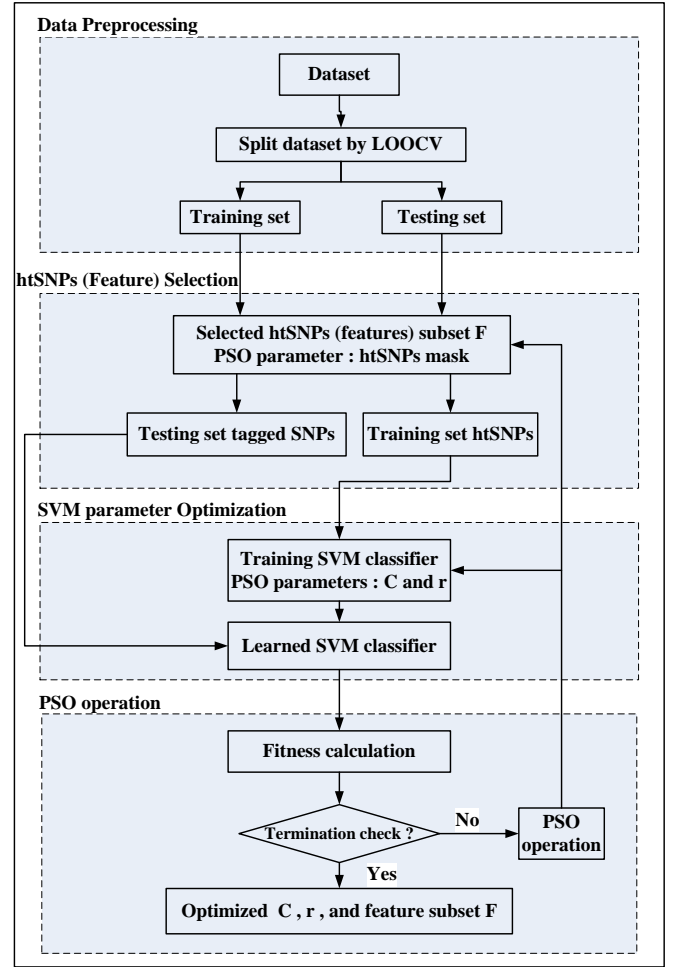


Figure 2 The flowchart of the proposed PSO-SVM model.

measurement mentioned above, details of the proposed hybrid PSO-SVM procedure are described as follows:

Procedure PSO-SVM hybrid model

1.) Data preparation

Given a dataset U is considered using the leave-one-out cross-validation process to split the data into training and testing sets. The training and testing sets are represented as U_{TR} and U_{TE} , respectively.

2.) PSO initialization and parameters setting

Set the PSO parameters including the number of iterations, number of particles, velocity limitation, particle dimension, disturbance rate. Generate initial particles comprised of the features mask, C and γ .

3.) Selected htSNPs (features) subset

Select input features for training set according to the feature mask which is represented in the particle from 2), then the features subset can be determined.

TABLE II. Results to compare PSO-SVM with SVM/STSA [4] and BPSO [5] on four real haplotype datasets.

Datasets (num of SNPs)		Prediction accuracy %											
		80	85	90	91	92	93	94	95	96	97	98	99
5q31 (103)	SVM/STSA	1	1	3	3	4	5	6	8	10	22	42	51
	BPSO	1	1	2	3	4	5	6	7	9	14	29	42
	PSO-SVM	1	1	2	2	3	4	5	6	7	10	23	36
TRPM8 (101)	SVM/STSA	1	1	2	5	5	6	7	8	10	15	15	24
	BPSO	1	1	2	5	5	6	7	8	9	13	14	22
	PSO-SVM	1	1	2	4	4	5	6	7	8	11	13	21
LPL (88)	SVM/STSA	2	3	4	10	13	20	25	30	35	39	42	47
	BPSO	2	3	6	9	12	16	18	21	25	28	31	37
	PSO-SVM	2	3	4	7	10	12	13	17	20	22	26	31

4.) SVM model training and testing

Based on the parameters C and γ which are represented in the particle, to train the SVM classifier on the training dataset, then the prediction accuracy for SVM on the testing dataset by LOOCV can be evaluated.

5.) Fitness calculation

For each particle, evaluate its fitness value by the prediction accuracy obtained from previous step. The optimal fitness value will be stored to provide feedback on the evolution process of PSO to find the increasing fitness of particle in the next generation.

6.) Termination check

When the maximal evolutionary epoch is reached, the program ends; otherwise, go to the next step.

7.) PSO operation

In the evolution process, discrete valued and continuous valued dimension of PSO with the disturbance operator may be applied to search for better solutions.

V. EXPERIMENTAL RESULTS AND COMPARISONS

To validate the performance of the developed hybrid approach, three public experimental SNP datasets [4] including 5q31, TRPM8 and LPL are used to compare the proposed approach with other previously published methods. When there are missing data exist in haplotype datasets, the GERBIL [4-5] program is used to resolve them. The chromosome 5q31 dataset was from the 616 kilobase region of human chromosome 5q31 and the SNPs were 103. The TRPM8 which consists of 101 SNPs was obtained from HapMap. The human lipoprotein lipase (LPL) gene was derived from the haplotypes of 71 individuals typed over 88 SNPs.

Our implementation platform was carried out on the Matlab 7.3, a mathematical development environment by extending the Libsvm which is originally designed by Chang and Lin [10]. The empirical evaluation was performed on Intel Pentium IV CPU running at 3.4GHz and 2 GB RAM. Through initial experiment, the parameter values of the PSO were set as follows. The swarm size is set to 200 particles. The searching

ranges of continuous type dimension parameters are: $C \in [10^{-2}, 10^4]$ and $\gamma \in [10^{-4}, 10^4]$. The discrete type particle for features mask, we set $[V_{\min}, V_{\max}] = [-6, 6]$, which yields a range of $[0.9975, 0.0025]$ using the sigmoid limiting transformation by Eq. (4). Both the cognition learning factor c_1 and the social learning factor c_2 are set to 2. The disturbance rate is 0.05, and the number of generation is 600. The inertia weight factor $w_{\min} = 0.4$ and $w_{\max} = 0.9$. The linearly decreasing inertia weight is set as Eq. (13), where i_{now} is the current iteration and i_{max} is the pre-defined maximum iteration.

$$w = w_{\max} - \frac{i_{\text{now}}}{i_{\text{max}}}(w_{\max} - w_{\min}) \quad (13)$$

To compare the proposed PSO-SVM approach with the SVM/STSA in [4] and BPSO in [5] on the three haplotype datasets by LOOCV experiments, the computational results of prediction accuracy according to the numbers of selected htSNPs are summarized in Table 2. As mentioned in [4], it is astonished that only one SNP for the 80% prediction accuracy in 5q31 and TRPM8 datasets can be achieved. In practice, if one guesses each SNP as 0, the prediction accuracy of 72.5% for 5q31 dataset and 79.3% for TRPM8 dataset would be obtained. Therefore, the appropriate selection of one htSNPs to correctly predict 80% on the rest of non-selected SNPs is reasonable. It is obvious that the proposed PSO-SVM hybrid model achieves higher prediction accuracy with fewer selected htSNPs in the three haplotype datasets. In general, the prediction accuracy is increased refers to the incremental selected htSNPs number. From Figure 3 to Figure 5 show that the numbers of selected htSNPs on haplotype datasets are proportional to the prediction accuracy and the PSO-SVM algorithm has very good performance for haplotype tagging SNPs selection problem in the three testing cases.

VI. CONCLUSION

In this paper, a hybrid PSO-SVM model that combines the particle swarm optimization (PSO) and support vector machine (SVM) with feature selection and parameter optimization is proposed to effectively solve for the haplotype tagging SNP selection problem. Several public datasets of different sizes are considered to compare the PSO-SVM with SVM/STSA and BPSO previously published methods. The experimental results show that the effectiveness of the proposed approach and the high prediction accuracy with the fewer number of haplotype tagging SNP can be obtained by the hybrid PSO-SVM system.

REFERENCES

- [1] K. Zhang, M. Deng, T. Chen, M. Waterman and F. Sun, "A dynamic programming algorithm for haplotype block partitioning," *Proc. Natl. Acad. Sci.*, vol. 99, pp. 7335-7339, 2002.
- [2] K. Zhang, F. Sun, S. Waterman and T. Chen, "Haplotype block partition with limited resources and applications to human chromosome 21 haplotype data," *Am. J. Hum. Genet.*, vol. 73, pp. 63-73, 2003.
- [3] H. Avi-Itzhak, X. Su, and F. de la Vega, "Selection of minimum subsets of single nucleotide polymorphisms to capture haplotype block diversity," In *Proceedings of Pacific Symposium on Biocomputing*, vol. 8, pp. 466-477, 2003.
- [4] He Jingwu and A. Zelikovsky, "Informative SNP Selection Methods Based on SNP Prediction," *IEEE Transactions on NanoBioscience*, Vol. 6, pp. 60-67, 2007.
- [5] Cheng-Hong Yang, Chang-Hsuan Ho and Li-Yeh Chuang, "Improved tag SNP selection using binary particle swarm optimization," *IEEE Congress on Evolutionary Computation (CEC 2008)*, pp. 854-860, 2008.
- [6] V.N. Vapnik, "The nature of statistical learning theory," New York: Springer-Verlag, 1995.
- [7] E. Halperin, G. Kimmel and R. Shamir, "Tag SNP selection in genotype data for maximizing SNP prediction accuracy," *Bioinformatics*, Vol. 21, pp. 195-203, 2005.
- [8] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, Piscataway, NJ, 1997, pp. 4104-4109.
- [9] J.C. Bean, "Genetics and random keys for sequencing and optimization," *ORSA J. Comput.*, Vol. 6, pp. 154-160, 1994.
- [10] C.C. Chang, and C.J. Lin, LIBSVM: a library for support vector machines, Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.

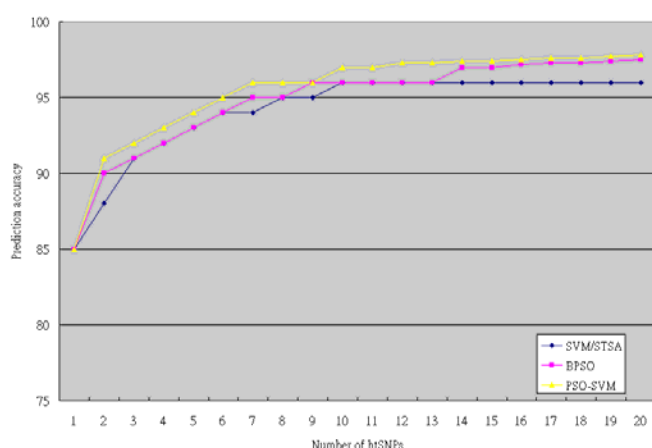


Figure 3 The comparison result of prediction accuracy associated with selected htSNPs on 5q31 datasets.

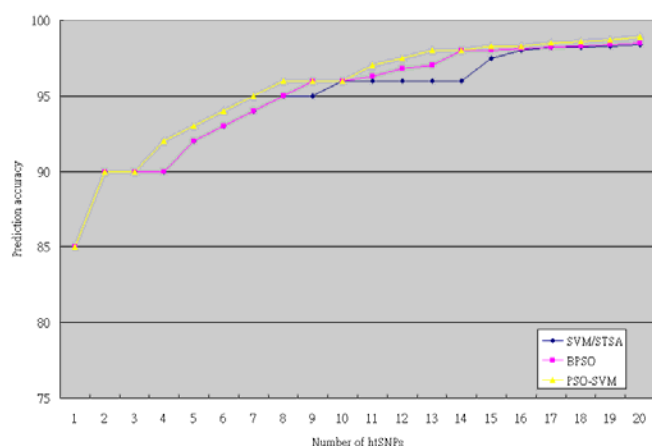


Figure 4 The comparison result of prediction accuracy associated with selected htSNPs on TRPM8 datasets.

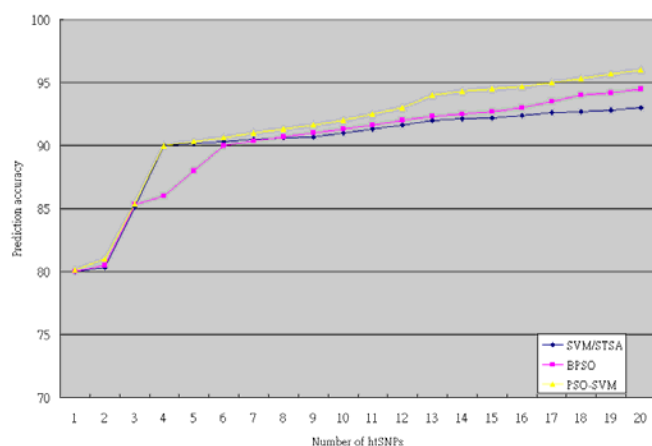


Figure 5 The comparison result of prediction accuracy associated with selected htSNPs on LPL datasets.

PAPR Reduction Technique for LTE SC-FDMA Systems Using Root-Raised Cosine Filter

Md. Masud Rana, Jinsang Kim and Won-Kyung Cho

Department of Electronics and Radio Engineering, Kyung Hee University
1 Seocheon, Kihung, Yongin, Gyeonggi, 449-701, Republic of Korea
Email: mamaraece28@yahoo.com

Abstract—Recently, mobile radio communications have developed rapidly due to the endless demand for broadband multimedia access and wireless connection anywhere, and any time. With the emergence of diverse fourth generation (4G) enabling technologies, signal processing has become ever increasingly important for small power, small chip resources, and efficient physical implementations of potential multimedia wireless communication systems. In this paper, we analytically derive the time and frequency domain single carrier-frequency division multiplexing (SC-FDMA) signals. Simulation results show that the SC-FDMA sub-carrier mapping scheme has a significantly lower peak-to average power ratio (PAPR) compared to orthogonal frequency division multiplexing (OFDMA). In addition, the interleaved FDMA (IFDMA) sub-carrier mapping scheme with root raised cosine filter reduced PAPR significantly than localized FDMA (LFDMA) and distributed (DFDMA) sub-carrier mapping scheme. As a result, it improves the mean power output from a battery driven terminal equipment and power amplifier efficiency.

Index Terms—CCDF, IFDMA, OFDMA, PAPR, root-raised cosine, SC-FDMA.

I. INTRODUCTION

The further increasing demand on high data rates in wireless communication systems has arisen in order to support broadband services. The third generation partnership project (3GPP) members started feasibility study on the enhancement of the universal terrestrial radio access (UTRA) in December 2004, to improve the mobile phone standard to cope with future requirements. This project was called long term evolution (LTE) [1].

LTE uses single carrier frequency division multiple access (SC-FDMA) for uplink transmission and orthogonal frequency division multiple access (OFDMA) for downlink transmission [6]. SC-FDMA is a promising technique for high data rate transmission that utilizes single carrier modulation and frequency domain equalization. Single carrier transmitter structure leads to keep the peak-to average power ratio (PAPR) as low as possible that is reduced the energy consumption. SC-FDMA has similar throughput performance and essentially the same overall complexity as OFDMA [3], [10], [12]. A highly efficient way to cope with the frequency selectivity of wideband channel is OFDMA. OFDMA is an effective technique for combating multipath fading and for high bit

rate transmission over mobile wireless channels. In OFDMA system, the entire channel is divided into many narrow sub-channels, which are transmitted in parallel, thereby increasing the symbol duration and reducing the intersymbol-interference (ISI) [4], [8]. Despite many benefits of OFDMA for high speed data rate services, it suffers from high envelope fluctuation in the time domain, leading to large PAPR. Because the high PAPR is detrimental to user mobile equipment (UE) terminals, SC-FDMA has drawn great attention as an attractive alternative to OFDMA for uplink data transmission. It can be regarded as DFT-spread OFDMA (DFTS-OFDM), where time domain data signals are transformed to frequency domain by a DFT before going through OFDMA modulation. The main benefit of DFTS-OFDM compared to OFDM scheme, is reduced variations in the instantaneous transmit power, implying the possibility for increased power-amplifier efficiency, low-complexity high-quality equalization in the frequency domain, and flexible bandwidth assignment [12].

In order to solve the high PAPR problem seen in the uplink of OFDMA, research is now addressing techniques such as a SC-FDMA. The most of the previous work related to 3GPP LTE uplink has been mainly focused on implementation problems in the physical layer [2], [5], [9], [13]. In [10], [12] proposed raised-cosine pulse shaping method that compares PAPR characteristics using the complementary cumulative distribution function (CCDF) for different subcarrier mapping.

PAPR reduction is of the most importance performance parameter in case of high amplitude signals subject to non linear power amplification. This situation more and more occurs due to the ever-growing demand in high spectral efficiency advanced mobile telecommunications systems implying multi dimensional waveforms considerations for which the PAPR is high. Pulse shaping is required for a single carrier system to bandlimit the transmit signal. This paper addresses a theoretical analysis of the PAPR reduction of LTE SC-FDMA systems when root-raised cosine (RRC) filter is used. RRC is used as the transmit and receive filter in a digital communication system to perform matched filtering. The combined response of two such filters is that of the raised-cosine filter. In this paper, we analytically derive the time and frequency domain SC-FDMA signals. Simulation results show

that the SC-FDMA has a significantly lower PAPR compared to OFDMA system. In addition, we comparing the three forms of SC-FDMA sub-carrier mapping scheme and find that the interleave FDMA (IFDMA) sub-carrier mapping with root raised cosine based pulse shapping method reduced PAPR significantly than localized FDMA (LFDMA) and DFDMA sub-carrier mapping scheme. As a results, improves the mean power output from a battery driven terminal equipment and power amplifier efficiency.

The rest of the paper is organized as follows. We describes the 3GPP LTE and LTE SC-FDMA system model in section II and III, respectively. In section IV, we describes the different SC-FDMA sub-carrier mapping scheme. In section V, we describes the PAPR reduction technique for LTE SC-FDMA systems. In section VI, we simulated and compare the proposed method with OFDMA for different sub-carrier mapping scheme. Finally, conclusions are made in section VII.

II. 3GPP LTE

The main purposes of the 3GPP LTE are substantially improved end-user throughputs, low latency, reduced user equipment (UE) complexity, high data rate, and significantly improved user experience with full mobility. First 3GPP LTE and LTE-advanced (LTE-A) specification is being finalized within 3GPP release 9 and release 10, respectively [1].

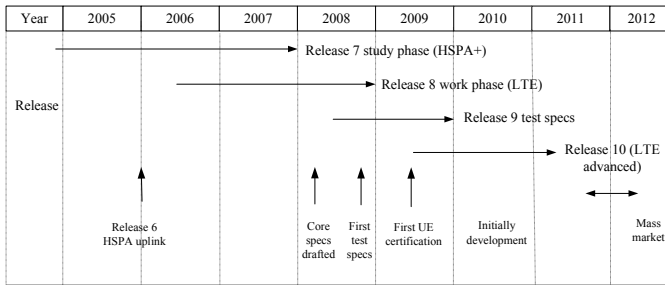


Fig. 1. LTE release timeline.

Specifically, the physical layer has become quite stable recently for a first implementation. LTE supports multipule input multiput output (MIMO) with one, two, four, and eight antenna elements at base station (BS) and mobile terminal. Both closed and open loop MIMO operation is possible. The target of LTE-A is to reach and surpass the international telecommunication union (ITU) requirements. One of the important LTE-A benefits is the ability to leverage advanced topology networks; optimized heterogeneous networks with a mix of macros with low power nodes such as picocells, femtocells, ensures user fairness, worldwide roaming, and new relay nodes [1].

In 3GPP LTE, the basic unit of a transmission scheme is a radio frame which is ten msec long. They are divided into ten sub-frames, each sub-frame one msec long. Each sub-frame is further divided into two slots, each of half msec duration. Fig. 2, shows the basic LTE generic frame structure [11]. The sub-carrier spacing in the frequency domain is 15 kHz. Twelve of these sub-carriers together (per slot) is called a resource block therefore one resource block is 180 kHz. Six resource

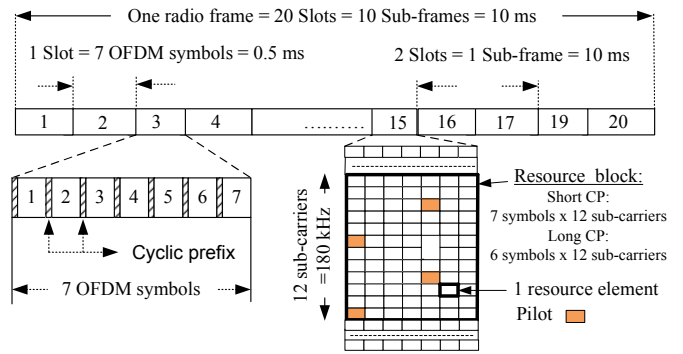


Fig. 2. LTE generic frame structure.

blocks fit in a carrier of 1.4 MHz and 100 resource blocks fit in a carrier of 20 MHz. Slots consist of either 6 or 7 OFDM symbols, depending on whether the normal or extended cyclic prefix (CP) is employed. The CP is added in front of each block. The details transmission scheme parameters of the 3GPP LTE system are shown in Table I [7]. LTE uses SC-FDMA scheme for the uplink transmissions and OFDMA in downlink transmission.

TABLE I
LTE SYSTEM PARAMETERS

Trans. bandwidth (MHz)	1.25	2.5	5	10	15	20
FFT size	128	256	512	1024	1536	2048
Occupied sub-carrier	76	151	301	601	901	1200
Sampling frequency (MHz)	1.92	3.84	7.68	15.36	23.04	30.72
No. of available PRBs	6	12	25	50	75	100
User plane latency (ms)	< 5					
PRB bandwidth (kHz)	180					
Frame duration (ms)	0.5					
Sub-carrier bandwidth (kHz)	15					
Coverage (km)	5-30					
Mobility (km/hr)	15-350					
Peak data rates (Mbits/s)	DL: 100, and UL: 50					
Antenna configuration	DL: 4x2, 2x2, 1x2, 1x1, and UL: 1x2, 1x1					
Spectrum efficiency	DL: 3-4 x HSDPA, and UL: 2-3 x HSUPA Rel.6					
Control plane latency (ms)	100 (idle to active), and 50 (dormant to active)					
Radio resource	DL: 3-4 fold higher than Rel.6					

III. LTE SC-FDAMA SYSTEM MODEL

The basic principle of a LTE SC-FDMA transmission system is shown in Fig. 3.

At the transmitter side, a baseband modulator transmits the binary input to a multilevel sequences of complex number $m_1(q)$ in one of several possible modulation formats including, quandary phase shift keying (QPSK), and 16 level-QAM. These modulated symbols are perform a N-point discrete Fourier transform (DFT) to produce a frequency domain representation [3]:

$$s_1(n) = \frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} m_1(q) e^{-j \frac{2\pi q n}{N}}, \quad (1)$$

where m_1 is the discrete symbols, q is the sample index, j is the imaginary unit, and $m_1(q)$ is the data symbol. The output

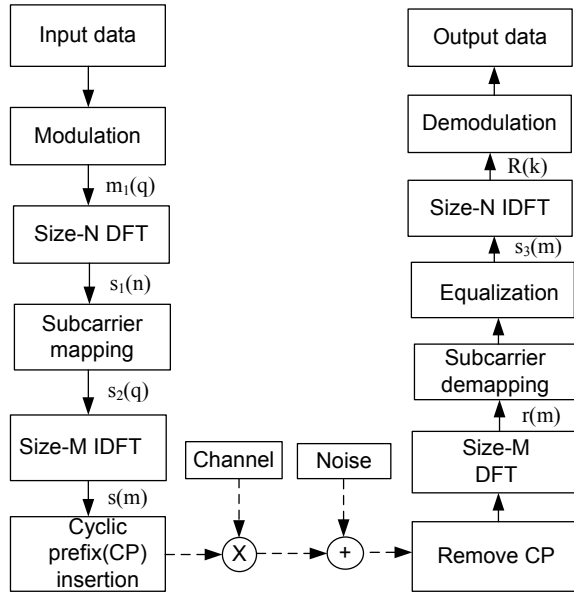


Fig. 3. LTE SC-FDMA transceiver system model [6].

of the DFT is then applied to consecutive inputs of a size- M inverse DFT ($M > N$) and where the unused inputs of the IDFT are set to zero. If they are equal ($M=N$), they simply cancel out and it becomes a conventional single user single carrier system with frequency domain equalization. However, if N is smaller than M and the remaining inputs to the IDFT are set to zero, the output of the IDFT will be a signal with 'single-carrier' properties, i.e. a signal with low power variations, and with a bandwidth that depends on N . The SC-FDMA is single-carrier, not single frequency. The data signal of each user consists of a lot of frequency. DFT of SC-FDMA is used to filter the frequency items and maps them into IDFT to reform single user waveform. This may justify the reduced peak-to-average power ratio (PAPR) experienced in the IDFT output. The details description of the sub-carrier mapping mode are in section IV. PAPR is a comparison of the peak power detected over a period of sample occurs over the same time period. The PAPR of the transmit signal is defined as [14]:

$$PAPR = \frac{\max_{0 \leq m < T} |s(m)|^2}{\frac{1}{TN} \int_0^{TN} |s(m)|^2 dm}, \quad (2)$$

where T is the symbol period of the transmitted signal $s(m)$. PAPR is best described by its statistical parameter, complementary cumulative distribution function (CCDF). CCDF measures the probability of signal PAPR exceeding certain threshold [12], [14]. To further reduce the power variations of the DFTS-OFDM signal, explicit spectrum shaping can be applied. Spectrum shaping is applied by multiplying the frequency samples with some spectrum-shaping function, e.g. a root-raised-cosine function (raised-cosine-shaped power spectrum). The IDFT module output is followed by a CP insertion that completes the digital stage of the signal flow. A CP is used to eliminate ISI and preserve the orthogonality of the

tones. Assume that the channel length of CP is larger than the channel delay spread [8].

The transmitted symbols propagating through the radio channel can be modeled as a circular convolution between the channel impulse response (CIR) and transmitted data blocks. At the receiver, the opposite set of the operation is performed. The CP samples are discarded and the remaining N samples are processed by the DFT to retrieve the complex constellation symbols transmitted over the orthogonal sub-channels. The received signals are de-mapped and equalizer is used to compensate for the radio channel frequency selectivity. After IDFT operation, the corresponding output is demodulated and soft or hard values of the corresponding bits are passed to the decoder.

IV. SC-FDMA SUB-CARRIER MAPPING SCHEME

There are two principal sub-carrier mapping modes: localized mode, and distribution mode. An example of SC-FDMA transmit symbols in the frequency domain for two user, three sub-carrier per user and six sub-carriers in total is illustrated in Fig. 4 [6].

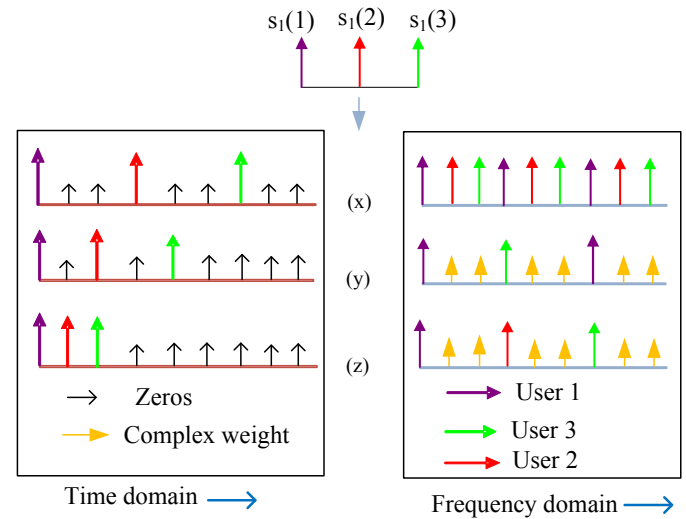


Fig. 4. Multiple access scheme of SC-FDMA: (x) IFDMA mode, (y) DFDMA mode, and (z) LFDMA.

In distributed sub-carrier mode, the outputs are allocated equally spaced sub-carrier, with zeros occupying the unused sub-carrier in between. While in localized sub-carrier mode, the outputs are confined to a continuous spectrum of sub-carrier [10], [12]. Except the above two modes, interleaved sub-carrier mapping mode of SC-FDMA (IFDMA) is another special sub-carrier mapping mode. The difference between DFDMA and IFDMA is that the outputs of IFDMA are allocated over the entire bandwidth, whereas the DFDMA's outputs are allocated every several sub-carriers. If there are more than one user in the system, different sub-carrier mapping modes give different sub-carrier allocation [10], [12]. In order to accommodate multiple access to the system and to preserve the constant envelope property of the signal, the

elements of transmitted signal $s(m)$ are mapped, by using the LFDMA or IFDMA sub-carrier mapping rule.

Here is output symbol calculation in IFDMA in time domain. The frequency samples after IFDMA sub-carrier mapping is [12], [16]:

$$s_2(q) = \begin{cases} s_1(q/C) = 1 & 0 \leq n \leq N-1 \\ 0 & \text{else} \end{cases}$$

where $q = Cn$, and $C = M/N$ and it is the bandwidth expansion factor of the symbol sequence. If $N = M/C$ and all terminals transmit N symbols per block, the system can handle C simultaneous transmissions without co-channel interference. After M -point IDFT operation ($M > N$), time domain signal can be described as follows. Let $m = Nc + q$, where $0 \leq c \leq C-1$ and $0 \leq q \leq N-1$. The time domain IFDMA transmitted signal can be express as [12], [16]:

$$\begin{aligned} s(m) &= \frac{1}{M} \sum_{q=0}^{M-1} s_2(q) e^{j \frac{2\pi m q}{M}} \\ &= \frac{1}{NC} \sum_{n=0}^{N-1} s_1(n) e^{j \frac{2\pi m n}{N}} \\ &= \frac{1}{C} \frac{1}{N} \sum_{n=0}^{N-1} s_1(n) e^{j \frac{2\pi (Nc+q)n}{N}} \\ &= \frac{1}{C} \left[\frac{1}{N} \sum_{n=0}^{N-1} s_1(n) e^{j \frac{2\pi q n}{N}} \right]. \end{aligned} \quad (3)$$

The square backed $[\cdot]$ of the above equation represent the N -point IDFT operation. The above equation can be rewritten as

$$s(m) = \frac{1}{C} m_1(q), \quad (4)$$

where $m_1(q) = \frac{1}{N} \sum_{n=0}^{N-1} s_1(n) e^{j \frac{2\pi q n}{N}}$ is the N -point IDFT operation. Therefore in case of IFDMA, every output symbol is simple a repeat of input symbol with a scaling factor of $1/C$ in time domain. When the sub-carrier frequency allocation starts from f th sub-carrier i.e. $q = Cn + f$ then the frequency samples after IFDMA sub-carrier mapping is

$$s_2(q) = \begin{cases} s_1(q/C - f) = 1 & 0 \leq n \leq N-1 \\ 0 & \text{else} \end{cases}$$

and the time domain transmitted signal can be express as

$$\begin{aligned} s(m) &= \frac{1}{C} \left[\frac{1}{N} \sum_{n=0}^{N-1} s_1(n) e^{j \frac{2\pi q n}{N}} \right] e^{j \frac{2\pi m f}{M}} \\ &= \frac{1}{C} e^{j \frac{2\pi m f}{M}} m_1(q). \end{aligned} \quad (5)$$

Thus, when the sub-carrier frequency allocation starts from f th instead of zero then there is an additional phase rotation of $e^{j \frac{2\pi m f}{M}}$.

Here is output symbol calculation in LFDMA sub-carrier mapping in time domain. After DFT operation the frequency domain sub-carrier mapping signal can be written as

$$s_2(q) = \begin{cases} s_1(l) = 1 & 0 \leq l \leq N-1 \\ 0 & N \leq l \leq M-1 \end{cases}$$

After IDFT operation, time domain signal can be described as follows. Let $m = Cq + c$, where $0 \leq c \leq N-1$ and $0 \leq q \leq C-1$. Then time domain transmitted signal can be represent as [12], [16]:

$$\begin{aligned} s(m) &= \frac{1}{M} \sum_{l=0}^{M-1} s_2(q) e^{j \frac{2\pi m l}{M}} \\ &= \frac{1}{C} \frac{1}{N} \sum_{l=0}^{N-1} s_1(l) e^{j \frac{2\pi (Cq+c)l}{CN}} \end{aligned} \quad (6)$$

if $c = 0$, then

$$\begin{aligned} s(m) &= \frac{1}{C} \left[\frac{1}{N} \sum_{l=0}^{N-1} s_1(l) e^{j \frac{2\pi q l}{N}} \right] \\ &= \frac{1}{C} m_1(q) \end{aligned} \quad (7)$$

Since $m_1(q) = \sum_{p=0}^{N-1} m_1(p) e^{-j \frac{2\pi l p}{N}}$, and if $c \neq 0$, then

$$\begin{aligned} s(m) &= \frac{1}{NC} \sum_{l=0}^{N-1} \left[\sum_{p=0}^{N-1} m_1(p) e^{-j \frac{2\pi l p}{N}} \right] e^{j \frac{2\pi (Cq+c)l}{CN}} \\ &= \frac{1}{NC} \sum_{l=0}^{N-1} \sum_{p=0}^{N-1} m_1(p) e^{j \frac{2\pi (q-p)l}{N+c/CN}} \\ &= \frac{1}{NC} \sum_{p=0}^{N-1} m_1(p) \frac{1 - e^{j \frac{2\pi (q-p)l}{N+c/CN}}}{1 - e^{j \frac{2\pi p l}{N+c/CN}}} \\ &= \frac{1}{NC} \sum_{p=0}^{N-1} m_1(p) \frac{1 - e^{j \frac{2\pi p l}{N+c/CN}}}{1 - e^{j \frac{2\pi (q-p)l}{N+c/CN}}} \\ &= \frac{1}{C} (1 - e^{j \frac{2\pi p l}{N+c/CN}}) \frac{1}{N} \sum_{n=0}^{N-1} \frac{m_1(p)}{1 - e^{j \frac{2\pi (q-p)l}{N+c/CN}}} \end{aligned} \quad (8)$$

So, the time domain LFDMA sub-carrier mapping signal has exact copies of input time signals with a scaling factor of $1/C$ in the N -multiple sample positions and in between values are sum of all the time input symbols in the input block with different complex-weighting (CW) [12].

Here is output symbol calculation in DFDMA in time domain. The frequency samples after DFDMA sub-carrier mapping is [12], [16]:

$$s_2(q) = \begin{cases} s_1(q/\tilde{C}) = 1 & 0 \leq n \leq N-1 \\ 0 & \text{else} \end{cases}$$

where $0 \leq c \leq C-1$, $q = Cn$, and $0 \leq \tilde{C} \leq C$. Let $m = Nq + c$ where $0 \leq c \leq C-1$ and $0 \leq q \leq N-1$. The time domain DFDMA transmitted signal can be express as [12], [16]:

$$\begin{aligned} s(m) &= \frac{1}{M} \sum_{l=0}^{M-1} s_2(q) e^{j \frac{2\pi m l}{M}} \\ &= \frac{1}{C} \frac{1}{N} \sum_{l=0}^{N-1} s_1(l) e^{j \frac{2\pi (Cq+c)l}{CN}} \tilde{C}_l \end{aligned} \quad (9)$$

if $c = 0$ and according to the previous procedure, we obtain

$$s(m) = \frac{1}{C} m_1(\tilde{C}q) \quad (10)$$

Since $m_1(q) = \sum_{p=0}^{N-1} m_1(p) e^{-j2\pi qp/N}$, if $c \neq 0$ and according to the previous procedure, we obtain

$$s(m) = \frac{1}{C} (1 - e^{j2\pi \tilde{C}c/C}) \frac{1}{N} \sum_{n=0}^{N-1} \frac{m_1(p)}{1 - e^{(j2\pi(\tilde{C}q-p)/N + \tilde{C}c/CN)}} \quad (11)$$

So, the time domain symbols of DFDMA sub-carrier mapping have the same structure as those of LFDMA sub-carrier mapping.

V. PAPR REDUCTION TECHNIQUE FOR LTE SC-FDMA SYSTEMS

In case of high amplitude signals subject to non linear power amplification, PAPR reduction is one of the most importance performance parameter. This situation more and more occur due to the ever-growing demand in high spectral efficiency telecommunications systems implying multi dimensional waveforms considerations for which the PAPR is high. In a single-carrier communication system, pulse shaping is required to bandlimit the signal and ensures it meets the spectrum mask. In this paper, a root raised cosine (RRC) filter is used to pulse shape the SC-FDMA signals. RRC is used as the transmit and receive filter in a digital communication system to perform matched filtering. The combined response of two such filters is that of the raised-cosine filter. The raised-cosine filter is used for pulse-shaping in digital modulation due to its ability to minimize intersymbol interference (ISI). Its name stems from the fact that the non-zero portion of the frequency spectrum of its simplest form $\beta = 1$ (called the roll-off factor) is a cosine function, 'raised' up to sit above the f (horizontal) axis. The RRC filter is characterized by two values; β , and T_s (the reciprocal of the symbol-rate). The impulse response of such a filter can be given as:

$$h(t) = \begin{cases} 1 - \beta + 4\beta/\pi, & t = 0 \\ \beta/\sqrt{2}[(1 + 2/\pi) \sin(\pi\beta/4) + (1 - 2/\pi) \cos(\pi\beta/4)], & t = \pm T_s/\beta 4 \\ \frac{\sin[\pi t/T_s(1-\beta)] + 4\beta t/T_s \cos[\pi t/T_s(1+\beta)]}{\pi t/T_s [1 - (4\beta t/T_s)^2]}, & \text{else} \end{cases}$$

It should be noted that unlike the raised-cosine filter, the impulse response is not zero at the intervals of $\pm T_s$. However, the combined transmit and receive filters form a raised-cosine filter which does have zero at the intervals of $\pm T_s$. Only in the case of $\beta = 0$, does the root raised-cosine have zeros at $\pm T_s$.

VI. PERFORMANCE ANALYSIS

The performance of the aforementioned PAPR reduction technique is explored by performing extensive computer simulations. All simulation parameters of the LTE SC-FDMA systems are summarized in Table II [6].

The CCDF of PAPR, which is the probability that PAPR is higher than a certain PAPR value PAPR_0 , is calculated by Monte Carlo simulation. We compare the PAPR value that is

TABLE II
THE SYSTEM PARAMETERS FOR SIMULATIONS

System parameters	Assumptions
System bandwidth	5MHz
Number of sub-carriers	512
Data block size	16
Roll of factor	0.0999999999
Oversampling factor	4
Number of iteration	10^4
Sub-carrier mapping schemes	DFDMA, IFDMA, LFDMA
Modulation data type	Q-PSK and 16-QAM
Spreading factor for IFDMA	32
Spreading factor for DFDMA	31

exceeded with probability less than 0.1 percentile PAPR. The PAPR calculation using various sub-carrier mapping schemes for SC-FDMA and OFDMA system is shown in Fig. 5. The modulation scheme used for the calculation of PAPR is QPSK. It can be seen that SC-FDMA sub-carrier mapping schemes

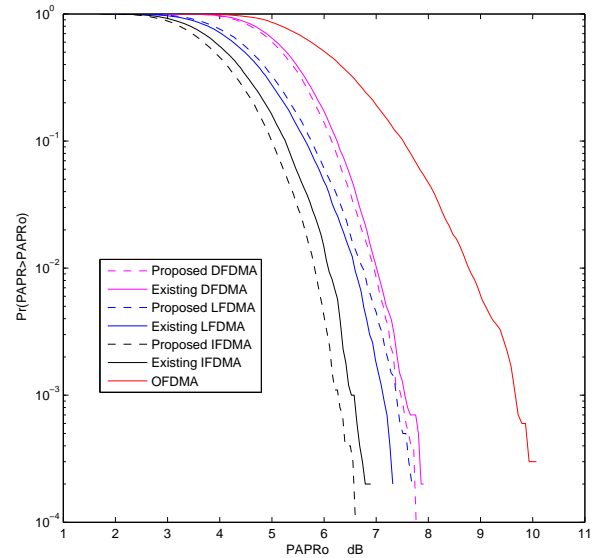


Fig. 5. Comparison of CCDF of PAPR for SC-FDMA with OFDMA using QPSK.

gives lower PAPR values as compared to OFDMA scheme. In addition, the root raised cosine pulse shaping method has lower PAPR than the case of existing pulse shaping method by more than 3dB. Due to the complex-weighting of LFDMA and DFDMA equation would increase the PAPR. Due to the phase rotation it is unlikely that the LFDMA samples will all add up constructively to produce a high output peak after pulse shaping. But it is shown that LFDMA has a lower PAPR than DFDMA when pulse shaping is applied. Another PAPR simulation using various sub-carrier mapping schemes for LTE SC-FDMA systems is shown in Fig. 6. The modulation scheme used for the calculation of PAPR is 16-QAM. It show that IFDMA has lowest value of PAPR at 7.8dB which is 6.7dB in case of QPSK as modulation technique. Finally, we conclude

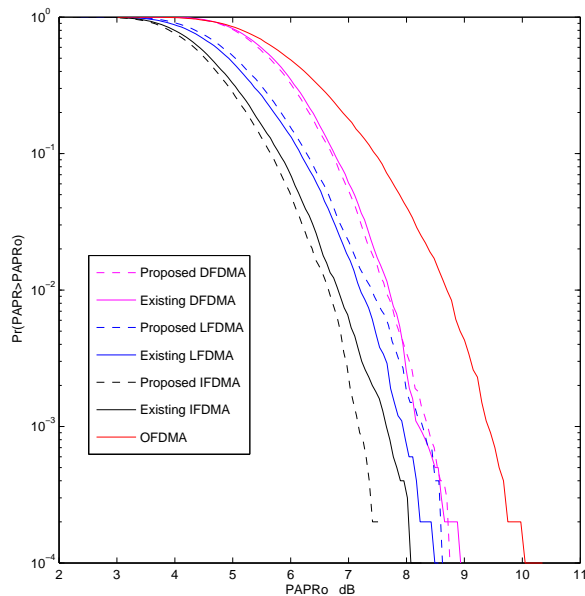


Fig. 6. Comparison of CCDF of PAPR for SC-FDMA with OFDMA using 16-QAM.

that the higher values of PAPR by using 16-QAM which is undesirable because they cause non linear distortions at the transmitter.

VII. CONCLUSIONS

The efficiency of a power amplifier is determined by the PAPR of the modulated signal. In this paper, we analysis different sub-carrier mapping scheme for LTE SC-FDMA systems. We derive the time and frequency domain signals of different sub-carrier mapping scheme, and numerically compare PAPR characteristics using CCDF of PAPR. We come to the conclusion that the IFDMA sub-carrier mapping with RRC pulse shaping method has lowest PAPR values compare to the other sub-carrier mapping methods. As a results, improves the mean power output from a battery driven terminal equipment and power amplifier efficiency. Therefore, SC-FDMA is attractive for uplink transmissions since it reduces the high PAPR seen with OFDMA.

ACKNOWLEDGMENT

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (201000171118).

REFERENCES

- [1] Q. Li, G. Li, W. Lee, M. Il. Lee, D. Clerckx, and Z. li, "MIMO techniques in WiMAX and LTE: a feature overview," *IEEE Commun. Magazine*, May 2010.
- [2] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, "3G evolution HSPA and LTE for mobile broadband," *Academic Press is an Imprint of Elsevier*, 2007.

- [3] B. Karakaya, H.Arsilan, and H. A. Cirpan, "Channel estimation for LTE uplink in high doppler spread," *Proc. WCNC*, pp. 1126-1130, April 2008.
- [4] J. Berkmann, C. Carbonelli, F.Dietrich, C. Drewes, and W. Xu, "On 3G LTE terminal implementation standard, algorithms, complexities and challenges," *Proc. Int. Con. on Wireless Communications and Mobile Computing*, pp. 970-975, August 2008.
- [5] A. Ancora, C. Bona, and D.T.M. Stock, "Down-sampled impulse response least-squares channel estimation for LTE OFDMA," *Proc. Int. Con. on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 293-296, April 2007.
- [6] M. M. Rana, M. S.Islam, and A. Z. Kouzani, "Peak to average power ratio analysis for LTE aystems," *Proc. Int. Con. on Communication Software and Networks*, pp. 516-520, February 2010.
- [7] A. Ancora, C. B. Meili, and D. T. Stock, "Down-sampled impulse response least- squares channel estimation for LTE OFDMA," *Proc. Int. Con. on Acoustics, Speech, and Signal Processing*, April 2007.
- [8] L. A. M. R. D. Temino, C. N. I Manchon, C. Rom, T. B. Sorensen, and P. Mogensen, "Iterative channel estimation with robust wiener filtering in LTE downlink," *Proc. Int. Con. on Vehicular Technology Conference*, pp. 1-5, September 2008.
- [9] J. Zyren, "Overview of the 3GPP long term evolution physical layer," *Dr. Wes McCoy, Technical Editor*, 2007.
- [10] H. G. Myung, J. Lim, and D. J. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30-38, September 2006.
- [11] M. Nouné and A. Nix, "Frequency-domain precoding for single carrier frequency- division multiple access," *IEEE Commun. Magazine*, vol. 48, no. 5, pp. 86-92, May 2010.
- [12] H.G. Myung, J. Lim, and D. J. Goodman, "Peak to average power ratio for single carrier FDMA signals," *Proc. PIMRC*, 2006.
- [13] S. Maruyama, S. Ogawa, and K.Chiba, "Mobile terminals toward LTE and requirements on device technologies," *Proc. Int. Con. on VLSI Circuits*, pp. 2-5, June 2007.
- [14] S. H. Han, and J. H. Lee, "An overview of peak to average power ratio reduction techniques for multicarrier transmission," *IEEE Transction on Wireless Communications*, April 2005.
- [15] H. G. Myung, "Introduction to single carrier FDMA," *Proc. Int. Con. on European Signal Processing (EUSIPCO)*, Poznan, Poland, September 2007.
- [16] A. Sohl, and A. Klein, "Comparison of localized, interleaved and block-interleaved FDMA in terms of pilot multiplexing and channel estimation," *Proc. Int. Con. on PIMRC*, 2007.

Survey of Routing Protocols and Channel Assignment protocols in Wireless Mesh Networks

Vivek M Rathod, Suhas J Manangi, Satish E, Saumya Hegde

National Institute of Technology Karnataka – Surathkal

Abstract: This paper is a survey on wireless mesh networks. Here we mention the basics of wireless mesh network, their purpose, channel assignment techniques and routing protocols. This survey is prepared towards helping those working on the relationship between channel assignment and routing protocols.

Keywords: Wireless Mesh Networks, Routing protocols, Channel Assignment, Multi Hop, Multi Radio.

I. INTRODUCTION

A wireless mesh network (WMN) ^[1] is a communication network made up of radio nodes organized in a mesh topology. The nodes which constitute the WMN are in adhoc mode so as to realize mesh topology.

Wireless mesh architecture is an effort towards providing high-bandwidth network over a specific coverage area. Wireless mesh architecture's infrastructure is, in effect, a router network minus the cabling between nodes. It's built of peer radio devices that don't have to be cabled to a wired port like traditional WLAN access points (AP) do. The traditional WLANs consist only of single hop end-to end connection (i.e., between the client and access point). In contrast, Mesh architecture sustains signal strength by breaking long distances into a series of shorter hops. Intermediate nodes not only boost the signal, but cooperatively make forwarding decisions based on their knowledge of the network, i.e. perform routing. Such architecture may with careful design provide high bandwidth, spectral efficiency, and economic advantage over the coverage area.

This paper is organized in the following sections.

1. Types of wireless mesh networks (network architectures).
2. Essential characteristics of WMN.
3. Components of WMN and their alternatives.
4. Routing purposes, problems and protocols.
5. Areas for research.

II. NETWORK ARCHITECTURE

The types of network structures being used for WMNs can be classified into three types in a very broad sense.

1. Client wireless mesh networks
2. Infrastructure wireless mesh networks
3. Hybrid wireless mesh networks ^[8].

A. Client WMN:

Client mesh networks or simply ad-hoc networks are formed by client devices with no supporting fixed infrastructure. Each node plays same role and participates in packet forwarding.

B. Infrastructure WMN:

In contrast to client WMN, infrastructure WMN consists of routers and client devices. The routers are interconnected via wireless links to form a multi-hop backhaul infrastructure. One or more routers are connected to the wired network and are called gateways of the WMN. Generally mesh router has two or more radio interfaces. One of which is an access interface for the clients to access the network. The second radio interface is a relay interface for forwarding and routing data packets. This is basically used for inter-router communication. Client devices associate themselves with the nearest mesh router to access the network. They don't participate in routing or relaying of packets. Therefore even if two clients are within the wireless range of each other, they cannot directly communicate. It has to happen through their respective routers.

C. Hybrid WMN^[8]:

This architecture is the combination of Infrastructure and client meshing; clients can access the network through mesh routers as well as by directly meshing with other mesh clients. A hybrid WMN is an extension to the Infrastructure WMN. In a hybrid WMN the clients not only connect to the wireless backhaul, but also serve as a gateway to for the clients which are located too far from the wireless mesh router. Therefore a hybrid WMN is more robust and more scalable than the previous two. A well-built hybrid WMN would enable fast, cheap and easy deployment of networks, leading to interesting applications such as emergency networks.

III. ESSENTIAL CHARACTERISTICS OF WMN

WMN has mainly following essential characteristics:

1. Multihop and Multi-pathing: multiple paths between two point in a WMN leads to the increase in bandwidth. This increase in the bandwidth is because the RTT (round trip time) for shorter paths (hops) is less than that of a single end to end path. Multiple packets can travel simultaneously

between two ends. Multi-pathing strips the data to be sent to a destination and sends these chunks via multiple paths, which increases the throughput significantly. It also provides robustness to paths, because there is always an alternative unless the destination itself is not connected to the network.

2. Self-healing, self-forming and self-organising: Since most of the nodes of the WMN are mobile, the WMN is always aware of its surroundings. It dynamically changes the routing paths based on the current state of the network. If a participating node quits, the network is reconfigured so as to keep the remaining nodes connected. Similarly the dynamic changes in the network must also take place based on the network traffic at different routes.
3. Compatibility and interoperability: The WMNs built on the IEEE 802.11 standards must be capable of supporting conventional Wi-Fi clients.
4. Cost Factor: WMNs can be very cost effective because we can build and configure a WMN with minimal existing resources. A WMN could provide an effective and good internet bandwidth to the group of users who share a single internet link.

IV. COMPONENTS OF WIRELESS MESH NETWORKS

A WMN consists of two types of wireless nodes. Mesh Routers and Mesh Clients. The Mesh Routers have improved computational, communication and power resources as compared to Mesh Clients. Mesh Routers are generally static and form the multi-hop backhaul network with optional access to other auxiliary networks. In addition, Mesh Routers are also typically equipped with multiple wireless network interfaces (IEEE 802.11^[3]) and are therefore able to establish high capacity connections. Mesh Clients are mobile devices, which take advantage of the existing communication infrastructure provided by the Mesh Routers.

A. 1-Radio VS Multi-Radio Approaches:

In 1-radio approach the participating nodes have only one radio each. Consider a network where both the clients and the mesh routers have only one radio, and then mesh routers would not be able to listen to the backhaul and the client simultaneously. Collisions would be very frequent. This will result in very low throughput. Thus one radio WMN is inferior to multi-radio infrastructure mesh networks in Multihop situations. In the case of 1 radio ad hoc mesh networks, available bandwidth is reduced by 50% with each hop: bandwidth available at the 3rd hop is 1/8 of the available capacity. However, while one-radio ad hoc mesh networks are unsuitable for Multihop situations, they are useful in one-hop situations for quickly establishing p2p communications. Conversely, 2-radio infrastructure meshes are ideal for Multihop situations with no restriction on the number of hops. Thus One radio mobile client mesh network combined with two or more

radio routers' backhaul support provides the best hybrid WMN; ubiquitous connectivity but with multiple levels of redundancy built in.

V. MULTI RADIO MESH ROUTER

There are mainly 4 types of Multi Radio Mesh Routers:

1. Single unit mesh router
2. split wireless router
3. Multi-Channel Multi-Transceiver single radio
4. low cost mesh router construction

A. Single Unit Mesh Router:

Single unit mesh router is a single package with multiple radios in it. All these radios' operate in non overlapping channels. Some of these could be used to relay packets between routers, while the others to provide connectivity to the clients or client adhoc network. Even though the radios operate in non-overlapping channels, the practical results have shown that there is a significant amount of interference between them due to the near-field effect, resulting in reduced throughput

B. Split Wireless Router:

Split mesh router is a network (wired) of two or more single radio routers. This design has gained motivation from the limitations of the single unit multi-radio routers. We refer to the single radio routers which are part of split router as nodes hereafter.

The commercially available single-radio routers often provide multiple interface technologies like the Ethernet, fiber or ATM. Two or more such units are connected via a backhaul using one of the available interface options like the Ethernet. Since the separation between these nodes is determined by the cable length forming the backhaul, the interference can be significantly reduced by increasing the distance between them. This is an effective solution for the interference due to near-field effect in the single unit mesh router.

Since our mesh router unit is a combination of 3-single radio routers, we need a software abstraction by which the assembly appears like a single unit to the network. Each single radio router must here be aware of the neighbors of the other two.

C. Multi-Channel Multi Transceiver Single Radio^[4]:

In this kind of routers, a radio includes multiple parallel RF front-end chips and baseband processing modules to support several simultaneous channels. On top of the physical layer, only one MAC layer module is needed to coordinate the functions of multiple channels. So far no multi-channel multi-transceiver MAC protocol has been proposed for WMNs.

D. Low Cost Mesh Router Construction

A low cost router can be set up using two USB or PCI radio cards on a low-end computer. But this would also require a MAC layer which supports multiple NICs simultaneously.

VI. CHANNEL ASSIGNMENT IN MULTI RADIO ROUTERS^{[2][4]}

In case of a two radio router network, there is no much flexibility, because one of the radios is used to communicate with other mesh routers on same channel and the other radio is used to communicate with clients.

But if we have more than two radios on each mesh router, we could use one for communicating with clients and the other radios could be intelligently assigned different channels so that they form channel diversified routes among mesh routers. Whether or not two routers are neighbors is decided by the channels assigned to them. The channel assignment can be made based on link quality and the topology.

There are various algorithms proposed for the channel assignment problem. They can be classified into two categories;

1. Interference-aware channel assignment(IACA)
2. Traffic-aware channel assignment (TACA)

Some of the algorithms are:

1. Identical channel assignment
2. Hybrid channel assignment
3. Centralized channel assignment
4. Maxflow based channel assignment routing (MCAR)
5. Topology and interference-aware channel assignment (TIC)

A. Identical Channel Assignment:

In this method first radio is assigned channel 1, second is assigned next non overlapping channel and so on. Though this preserves connectivity, this method in no way makes any effort in reducing interference.

B. Hybrid Channel Assignment:

In this strategy some radios are statically assigned channels while other radios are assigned channels dynamically.

C. Centralized Channel Assignment:

In this method the links are visited in some order and a common channel is assigned to interfaces on both the ends. If all the interfaces of the end node are already assigned a channel and they don't share any common channel, then it is necessary to replace one on the channel assignments. This ends up in a recursive channel assignment procedure. The visit can be in the decreasing order of the number of links falling in the interference range and the least used channel in that range is selected (interference aware). It could also be based on the estimated link flow rates(traffic aware).The algorithm might then visit all the links in decreasing order of expected link flow rate and select the channel which minimizes the sum of expected flow rates of all the links in the interference region that are assigned the same channel.

D. Maxflow Based Channel Assignment Routing:

MCAR is an improvement over centralized channel assignment algorithm. The interdependence among channel assignments across the whole network is taken into account by first identifying the groups of links that need to be assigned the same channel in order for the number of different channels on every router not to exceed the number of radios. Then, the actual channel assignment stage exploits the result of the first stage to assign channel in such a way that no replacement of previously assignments are necessary.

E. Topology and Interference aware Channel Assignment:

This algorithm undergoes two phases. One is Topology discovery and the other is channel selection.

Topology discovery: Prior to the channel assignment the topology is discovered. Topology discovery for every router is the identification of band-specific set of neighboring routers and the measurement of quality of link to each of these neighbors. Each router tunes itself to various channels on which band topology is to be discovered. This activity is co-ordinated by the channel management server. The link quality is measured by ETT (estimated transmit time).

Channel selection: Dijkstra's shortest path algorithm is used in TIC to discover frequency-diversified routes between the gateway and routers. The interference between mesh links is generated using conflict-graph model. For generating the above model interfering mesh links have to be identified in the first place. Thus the data generated in the first phase (topology discover) can be used to construct conflict graph. Thus the interfering links are assigned non-overlapping channels.

Cross-layer work: In most of the situation the throughput of configured WMN depends on both the channel assignment and routing algorithm chosen. So there is a lot of research in developing the cross-layer protocols which deals with the channel assignment and routing jointly.

VII. Routing Protocols^[9]

Routing protocols lie at the heart of designing a WMN network. They, in simple terms, specify the relay routes for packets in the network. Most of the protocols neglect the traffic between the mesh nodes and only consider the traffic between the nodes and the internet.

Network Asymmetry: This is the situation in which forward direction of a network is significantly different from the reverse direction in terms of bandwidth, loss rate, and latency. Forward path routing protocols are effective in routing of the packets from the mesh nodes to the gateway of the WMN, backward routing protocols are effective in routing the packets from the internet to the mesh nodes.

Some of the most popular protocols being used are AODV and OLSR.

A. Ad-hoc On Demand Distance Vector Routing(AODV):
The Ad hoc On Demand Distance Vector (AODV) routing algorithm is a routing protocol designed for ad hoc mobile networks. AODV is capable of both unicast and multicast routing. It is an on demand algorithm, meaning that it builds routes between nodes only as desired by source nodes. It maintains these routes as long as they are needed by the sources. Additionally, AODV forms trees which connect multicast group members. The trees are composed of the group members and the nodes needed to connect the members. AODV uses sequence numbers to ensure the freshness of routes. It is loop-free, self-starting, and scales to large numbers of mobile nodes.

AODV builds routes using a route request / route reply query cycle. When a source node desires a route to a destination for which it does not already have a route, it broadcasts a route request (RREQ) packet across the network. Nodes receiving this packet update their information for the source node and set up backwards pointers to the source node in the route tables. In addition to the source node's IP address, current sequence number, and broadcast ID, the RREQ also contains the most recent sequence number for the destination of which the source node is aware. A node receiving the RREQ may send a route reply (RREP) if it is either the destination or if it has a route to the destination with corresponding sequence number greater than or equal to that contained in the RREQ. If this is the case, then it unicast a RREP back to the source. Otherwise, it rebroadcasts the RREQ. Nodes keep track of the RREQ's source IP address and broadcast ID. If they receive a RREQ which they have already processed, they discard the RREQ and do not forward it. If the RREP propagates back to the source, then nodes set up forward pointers to the destination. Once the source node receives the RREP, it may begin to forward data packets to the destination. If the source later receives a RREP containing a greater sequence number or contains the same sequence number with a smaller hop count, it may update its routing information for that destination and begin using the better route.

As long as the route remains active, it will continue to be maintained. A route is considered active as long as there are data packets periodically travelling from the source to the destination along that path. Once the source stops sending data packets, the links will time out and eventually be deleted from the intermediate node routing tables. If a link break occurs while the route is active, the node upstream of the break propagates a route error (RERR) message to the source node to inform it of the now unreachable destination(s). After receiving the RERR, if the source node still desires the route, it can reinitiate route discovery.

Advantage and disadvantages:

The main advantage of this protocol is that routes are established on demand and destination sequence numbers are used to find the latest route to the destination. The connection setup delay is less. One of the disadvantages of

this protocol is that intermediate nodes can lead to inconsistent routes if the source sequence number is very old and the intermediate nodes have a higher but not the latest destination sequence number, thereby having stale entries. Also multiple RouteReply packets in response to a single RouteRequest packet can lead to heavy control overhead. Another disadvantage of AODV is that the periodic beaconing leads to unnecessary bandwidth consumption.

B. Optimized Link State Routing Protocol(OLSR):

It is a proactive protocol. The Optimized Link State Routing Protocol (OLSR) ^{[1] [5]} is a proactive routing protocol. Every node sends periodically broadcast "Hello"-messages with information to specific nodes in the network to exchange neighborhood information. The information includes the nodes IP, sequence number and a list of the distance information of the nodes neighbors. After receiving this information a node builds itself a routing table. Now the node can calculate with the shortest path algorithm the route to every node he wants to communicate. When a node receives an information packet with the same sequence number twice he is going to discard it. In these routing tables he stores the information of the route to each node in the network. The information is only updated:

1. A change in the neighborhood is detected.
2. A route to any destination is expired.
3. A better (shorter) route is detected for a destination.

The difference from OLSR to LSR (Links State Protocol) is that OLSR relies on multi-point relays (MPR). MPR is a node which is selected by its direct neighbor (one hop). The first idea of multipoint relays is to minimize the flooding of broadcast messages in the network. An information packet should not be sent twice in the same region of the network. MPR helps to optimize and reduce that problem. Each node informs its direct neighbors (one hop) about its MPR set in the "Hello"-messages. After receiving such a "Hello"-message, each node records the nodes MPR Selector that selects it as one of their MPRs. The second idea is that the size of the hello messages is reduced. It includes only the neighbors that select node N2 as one of their MPR nodes. In this way partial topology information is propagated. Node N2 can be reached only from its MPR selectors.

Advantages:

1. Minimal latency.
2. Ideal in high density and large networks.
3. OLSR achieves more efficiency than classic LS algorithms when networks are dense.
4. OLSR avoids the extra work of "finding" the destination by retaining a routing entry for each destination all the time, thus providing low single-packet transmission latency.
5. OLSR can easily be extended to QoS monitoring by including bandwidth and channel quality

information in link state entries. Thus, the quality of the path (e.g., bandwidth, delay) is known prior to call setup.

Disadvantages:

1. When the network is sparse, every neighbor of a node becomes a multipoint relay. The OLSR then reduces to a pure LS protocol.
2. High control overhead (reduced by MPR usage) .
3. Higher computation.
4. Storage.
5. Implementation complexity.

C. Backward Routing Protocol:

The above mentioned protocols are designed to route in the direction from the mesh nodes to the internet. However most services generate asymmetric traffic and the amount of the downstream from the servers in the internet to the mesh nodes far exceeds the upstream. Therefore some routing protocols are proposed which take care of this backward traffic. Backward path routing is more involved than routing on forward path because the data addressed at any host in the internet only needs to be forwarded to the gateway, while the backward routing protocol needs to address each node individually. There are three main families of backward routing protocols; reactive hop-by-hop routing, proactive hop-by-hop routing, and proactive source routing.

D. AODV-CGA:

This extended AODV protocol allows the use of multiple gateways to the internet. It shares most of the mechanisms with the well-known AODV protocol. The addition to the existing AODV which is made is that, all gateways are connected to a dedicated router that acts as a proxy to the internet. This router has two tasks:

1. On the forward path, it sends route on behalf of hosts in the internet;
2. On the backward path, it initiates route requests for nodes in the wireless mesh network.

E. Proactive Field-based Routing (PFR):

Wireless mesh nodes periodically exchange beacons. These beacons contain a list of all known destinations with their respective field value. When a new destination appears, it announces its presence with beacons to its neighbors in order to establish a field. With this mechanism, a field on the network is constructed for every destination. This field assigns a value to every node in the network; the destination bears the maximum value. Packets are then routed along the steepest gradient towards the destination.

Advantages and disadvantages:

This protocol ensures loop freedom. This protocol enables nodes to consider multiple routes to the destination. This protocol also has a drawback. Since it proactively maintains all routes, it incurs communication overhead even if the traffic is too low.

F. Gateway Source Routing (GSR):

In this protocol forward path information from the packets that arrive at the gateway is reused. In the routing header of every packet, the intermediate hops from the mesh node to the gateway are recorded. These paths are then stored in the gateways. To route packets to a mesh node, the mesh gateway inverts the recorded forward path and copies it to the packet header. The gateway then sends the packet to the first node of the backward path. Each node updates the path in the header by removing its entry and forward the packet to the given next hop until the packets reaches the destination.

GSR requires that a packet towards a host in the internet is first sent by a mesh node in order to establish the backward path. This should not be a problem when we assume that majority of communication is initiated by the mesh nodes.

If a mesh node has to act as a server, a dedicated addressing mechanism will have to be used.

G. Hierarchical Routing Protocols:

Some of the protocols which differ from the above are hierarchical routing and geography based routing.

In hierarchical routing, a certain self-organization scheme is employed to group network nodes into clusters. Each cluster has one or more cluster heads. Nodes in a cluster can be one or more hops away from the cluster head. Since connectivity between clusters is needed, some nodes can communicate with more than one cluster and work as a gateway. When the node density is high, hierarchical routing protocols tend to achieve much better performance because of less overhead, shorter average routing path, and quicker set-up procedure of routing path. However, the complexity of maintaining the hierarchy may compromise the performance of the routing protocol.

In WMNs, a mesh client must avoid being a cluster head because it can become a bottleneck due to its limited capacity.

H. Geographical Base Routing:

Compared to topology-based routing schemes, geographic routing schemes forward packets by only using the position information of nodes in the vicinity and the destination node. Thus, topology change has less impact on the geographic routing than the other routing protocols. Early geographic routing algorithms are a type of single-path greedy routing schemes in which the packet forwarding decision is made based on the location information of the current forwarding node, its neighbors, and the destination node. However, all greedy routing algorithms have a common problem, i.e., delivery is not guaranteed even if a path exists between source and destination. In order to guarantee delivery, planar-graph-based geographic routing algorithms have been proposed recently. However, these algorithms usually have much higher communication overhead than the single-path greedy routing algorithms.

VIII. CONCLUSIONS

Here we mentioned the basics of wireless mesh network, their purpose, various techniques involved and the area of the research in wireless mesh networks. Further research can be done studying the relationship between channel assignment techniques and routing protocol. These two areas will influence one another and an efficient combination can be possibly found.

IX. REFERENCES

- [1]. Optimization of routing algorithm in wireless mesh networks. Gupta B.K, Acharya B.M, Mishra M.K. NaBIC 2009
- [2]. Distributed channel assignment for multi-radio wireless mesh networks. Makram S>A, Gunes M. ISCC 2008
- [3]. Performance Analysis of IEEE802.11 Wireless Mesh Networks. Ye Yan, Hua Cai, Seung-Woo Seo. ICC 2008
- [4]. Routing metrics for multi-radio wireless mesh networks. Guerin J, Portmann M, Pirzada A. ATNAC 2007
- [5]. Routing Packets into Wireless Mesh Networks. Baumann R, Heimlicher S, Lenders V, May M. WiMOB 2007
- [6]. Neighbor selection technique for multi hop wireless mesh networks. Coll B, Gozalvez J.LCN 2009
- [7]. A New multi channel MAC protocol combined with on demand routing for wireless mesh networks. Guojun Shui, Shuqun Shen. ICCSSE 2008
- [8]. Hybrid routing with periodic updates in wireless mesh networks. Damle A, Rajan D, Faccin S.M. WCNC 2006
- [9]. Multipath routing algorithm based on traffic prediction in wireless mesh networks. Li Zhi yuan, Wang Ru chuan, Bi Jun lei. ICNC 2009

An Approach For Designing Distributed Real Time Database

Dr. Dhuha Basheer Abdullah

Computer Sciences Dept./Computers Sciences and
Mathematics College /Mosul University
Mosul- Iraq

Ammar Taher Yaseen

Computer Sciences Dept./Computers Sciences and
Mathematics College /Mosul University
Mosul- Iraq

Abstract- A distributed Real Time database system is a transaction processing system that is designed to handle workloads where transactions have service deadlines. The emphasis here is on satisfying the timing constraint of transactions (meet these deadlines, that is to process transactions before their deadlines expire) and investigating the distributed databases. This paper produces a proposed system named ADRTDBS.

In this work a prototype of client/server module and server/server module for distributed real time database has been designed. Server gets the data from direct user or a group of clients connected with it, analyze the request; and broad updating to all servers using 2PC (Two Phase Commit) and executing the demand by using 2PL (Two Phase Locking). The proposed model does not concern with data only, but provide a synchronize replication, so the updating on any server is not saved unless broadening the updating on all servers by using 2PC, and 2PL protocols. The database on this proposed system is homogenous and depend on full replication to satisfy real time requirements.

The transactions have been scheduled on the server by using a proposed algorithm named EDTDF (Earliest Data or Transaction Deadline First). This algorithm works to execute transactions that have smallest deadline at the beginning, either this deadline specific to the data or to the transaction itself. Implementing this algorithm helps to execute greater rate of transactions before their deadlines.

In this work two measures of performance for this system (proposed model) were been conducted; first, computing the Miss Ratio (rate of no. of executing transactions that miss their deadline); second, computing the CPU utilization (CPU utilization rate), by executing a set of transactions in many sessions.

Keywords: *real time, databases, distributed, replication, Scheduling*

I. INTRODUCTION

According to the definition provided by Coulouris, Dollimore & Kindberg [4], a distributed system consists of a set of autonomous processing elements that are connected via a communication network and interact via message passing.

A database is a structured set of data maintained by a database management system (DBMS) that interfaces with a set of applications or clients that access and modify the data. In a distributed database system, the data is distributed among autonomous DBMS instances (nodes or sites) that communicate via a network. The nodes, potentially along with a central

coordinator, are collectively referred to as a distributed database management system (DDBMS) [1,7,8,15].

In a distributed database, replication of data objects (The term object is used for the unit of replication; this could just as well be a table in a relational database as an object) is often used to improve fault tolerance and availability in the system by maintaining several copies of data objects and placing those copies close to the clients that want to use them [19].

In a real-time system (RTS), the value of a performed task depends not only on its functional correctness, but also on the time at which it is produced. For example, when an autonomous vehicle detects an obstacle in its intended path, it is crucial that it changes its path before a collision occurs. Real-time systems are often embedded, meaning that they are a part of (and interact heavily with) a physical environment. Typically, embedded systems use specific-purpose rather than general-purpose computers, such as in the embedded system controlling fuel injection in a car engine [6,20].

It is paramount that real-time systems have predictable, bounded and sufficiently low requirements on resources such as memory, network bandwidth and processor execution time, since failures due to unpredictable behavior and/or over consumption of available resources may cause unacceptable damage to humans or equipment. Real-time systems also need to be highly and predictably available, meaning that when a request is made to the system, it can be guaranteed that the system is available to service that request within a predictable and bounded time.

A distributed real-time system (DRTS) combines characteristics of distributed and real-time systems. This means that in such a system, issues related to distribution (such as execution of distributed algorithms and network communication) must be addressed with real-time requirements in mind.

Real-time database systems (RTDBS) are often used to manage data in real-time systems, since traditional databases cannot meet the timeliness and predictability requirements of a RTS. As many embedded applications with real-time requirements are inherently distributed, RTDBS are often distributed over a set of autonomous nodes, creating a need for distributed real-time database systems (DRTDBS) [10,14,16].

- Replication in DRTDBS

Data replication can be used to increase availability, predictability, and reliability of transaction processing in DRTDBS. Common replication approaches for DRTDBS use either a primary copy to deterministically apply updates to replicated data, or use distributed concurrency control and distributed commit protocols.

The distributed algorithms required to implement, e.g., distributed locking (to ensure serializability) and distributed commit (to ensure mutual consistency and durability) are hard to make predictable and sufficiently efficient due to their reliance on correct message delivery. Furthermore, depending on the replication approach a transaction may be forced to either wait or roll back and restart due to concurrent execution of transactions on remote nodes. Such behavior is problematic in real-time systems, since potential blocking times and rollbacks must be considered when determining worst-case execution times of transactions. For this reason, optimistic replication approaches, where transactions are allowed to execute as if no concurrent transactions exist, are more suitable than pessimistic replication approaches in real-time databases. Optimistic replication increase the availability, predictability and efficiency of transaction execution at the cost of transaction conflicts that must be resolved [2,9].

II. RATED WORK

In 1994 the researcher Nandt Subakr and others discuss the ways in which the Commit Protocol that could be adapted to the environmental sensitivity of the cases required for real-time. This protocol depends on the strategies and the installation optimistic on local compensation [13].

In 1994 also provided a researcher Victor Fiy and other researchers produce the basic rules to support the necessary qualities to the environment of the account distributed to RT, which is the modalities of distribution of time. It provides general concepts to clarify the application of the expansion of CORBA [18].

In 1998 the researcher Krayas Shihabi and others discuss the experience to implement the 2-Server have the same DBMS are linked through the Internet. Focusing on the intelligence linking the researchers explained how the firm Optimal query plan may choose the most expensive mistake. This takes precedence over the lack of knowledge of the operational environment [12].

In 2003 the researcher Yuan Wei and others discuss produce a study on the extraction using real-time updating of data and strategies on demand in DRTDB and the definition of certain laws to choose the best policy of modernization. Based on these laws, the researchers suggested an algorithm to derive

the updated data, the derivation policy of modernization practical data sets automatically [17].

In 2005 the researchers Broheedi Marcos and steen Andler illustrate how to bring forward the requirements in the DARTDBS. It is possible to use a model requirements of the modalities of information with RT[3].

In 2006 also provided a researcher Benoi Ravindran and others Where they distributed scheduling algorithm Call CUA. The parameters indicated it would satisfy for Thread time when there is failure. Algorithm is the Best-Effort and the Thread of the highest importance when they arrive at any time be the possibility of implementing a very high [11].

In 2008 the researcher Alexander Zharkov discuss how to use the material offers Materialized Views in DRTDBMSs. The researcher offers an algorithm for building dynamic and evaluation of the material cost. President difference this algorithm from its predecessors is taken into consideration the characteristics of time Temporal Properties of relations president and data processing [21].

III. CONTRIBUTIONS

ADRTDBS is a real time distributed database management system prototype that is designed to support distributed transactions processing with timing constraints.

The ADRTDBS offers many contributions listed below:

- *Database in main memory:* disk access should be minimize in a RTS, since reading from disk is both unpredictable and orders of magnitude slower than memory access. ADRTDBS is built to keep the entire database resident in main memory.
- *Full Replication:* times for network messages are unpredictable, and accessing data on remote nodes in much slower than local access. So ADRTDBS employs a full replication scheme which ensures that local replicas exist for all objects accessed by a transaction removing the need for remote object access.
- *ADRTDBS Design and Implementation:*
 - A structure to add support of executing real time transactions in distributed environment.
 - Providing scheduling algorithm named EDTDF Earliest Data or Transaction Deadline First for Transaction execution.
 - Produce an approach for concurrency control by using 2PL (Two Phase Locking) protocol to managing concurrent execution of transactions.
 - Execute data shipping and transaction shipping.
 - Provide synchronize replication and synchronization updating by using 2PC (Two Phase Commit).
 - Provide backup and recovery approach to process failure.

IV. THE PROPOSED SYSTEM

Given the important developments in computer and software industry databases and the increasing use in different areas of life (such as the management of banks, libraries, companies, factories ... etc.) and because of its great importance in a systematic compilation of data and processing, updating and retrieval with pinpoint accuracy, speed and the urgent need to provide such techniques in our country to keep pace with this development software tremendous invaded the whole world, this system was built to be a first step in the application of modern techniques and contemporary distributed database environment in real time. It was named Approach for designing Distributed Real Time Database System (ADRTDBS). The system ADRTDBS deals with Homogeneous distributed databases (i.e. all computers linked to the network is made of the same company (Pentium III) and contains the same version of the operating system (Windows XP) as well as containing the same version of the database management system (DBMS Oracle 9i), the same version of the program interfaces (Developer 6i). Has the capacity to implement Soft Real Time Transactions.

V. SYSTEM ARCHITECTURE

The system architecture consists of the following structure shown in the figure (1):

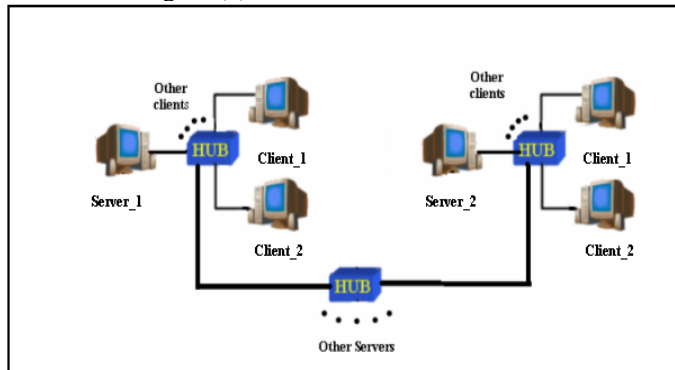


Figure (1) Architecture of ADRTDBS System

It contains two computers working as server having same database, and two computers working as clients connected with each server. Connection between computers is via HUB. The database resident in each server connected with the network and having same data and structure (replicas). The clients contain interfaces that making connection with servers and retrieve updating database. Any transaction will be executed on the database will have one of two cases: either local transaction, the implementation is the only current computer server. Or

Global transaction, the implementation to all servers linked in the network.

VI. SYSTEM MODEL

• Database Model

A real time distributed system consists of two autonomous computers system (sites) connected via a communication network. Each site maintains a copy of database. In order for transactions to be applied consistently to all replicas and give a result within deadline time, a prototype units runs at each site. Also this prototype architecture gives the distributed nature and the increased communication burden of such a database system.

The smallest unit of data accessible to the user is called data object. In this distributed database system with replicated data objects a logical data object is represented by a set of one or more replicated physical data object. The database is fully replicated at all sites. The database consists of two types of data objects: temporal and non-temporal. Temporal data object are used to record the state of the object in the external environment which its value changes frequently with time.

- Shipping Approaches

Two approaches for processing transactions in a ADRTDBS system: query shipping and data shipping.

• Data Shipping

In the data shipping approach, a transaction initiated by a client will be processed at the client. While the transaction is processing, the client sends data requests, which are required by the transaction, to the database server. The server responds to the requests by sending the required data objects to the client. The processing of the transaction will be completed at the client.

• Query Shipping

In the query shipping approach, the client sends queries to the database server for the transaction, instead of data requests. Once the server receives a query, it processes the query and sends the results back to the client. In the query shipping approach, the communication cost and the buffer space required at the client side are smaller than that in the data shipping approach. Also, the query shipping approach provides a relatively easy migration path from an existing single-site system to the client-server environment since the database engine can have a process structure similar to that of a single-site database system. On the other hand, the data shipping approach can off-load functionality from the server to the clients. This may improve the scalability of the system and balance the workload in the system. Figure (2) illustrates flowcharts for query shipping from the point of view for server and client.

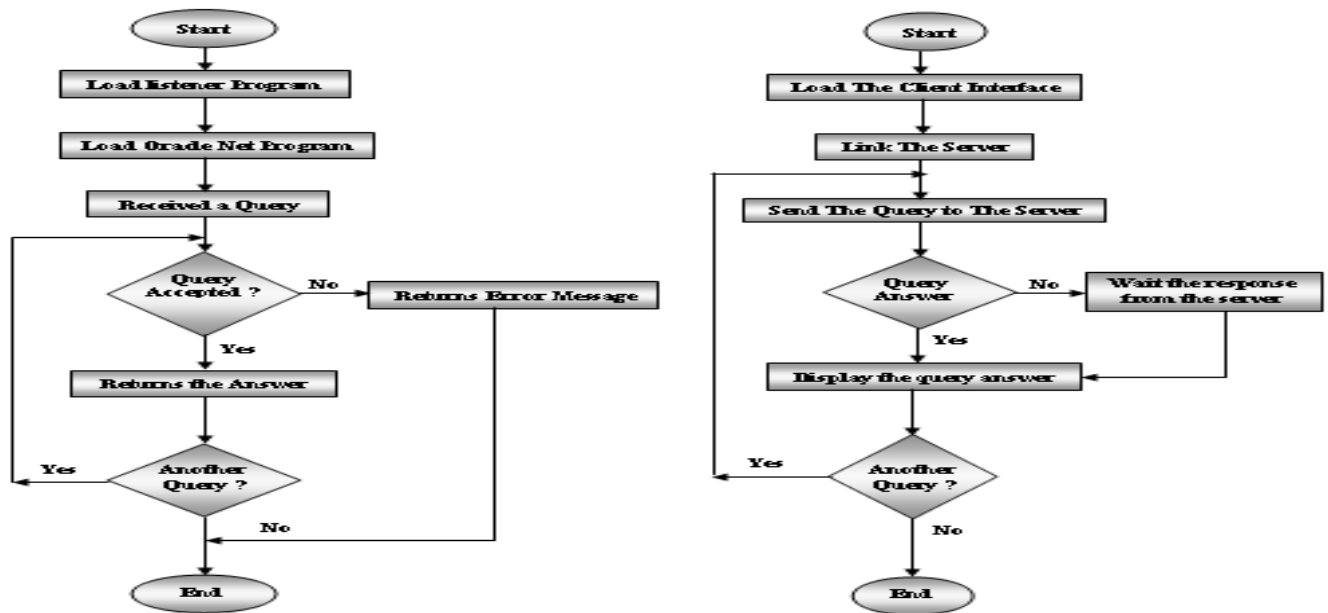


Figure (2) : a- Query Shipping in Server, b- Query Shipping in Client

• Transaction Model

A transaction is a sequence of operations that takes database from a consistent state to another consistent state. Two types of transactions are used in this proposed system: query transactions and update transactions. Query transactions consist only of read operations that access data object and return their values to the user. Thus, query transactions do not modify the database state. Update transactions consist of both read and write operations.

A transaction T_i in this proposed system characterized by the following attributes: $T_i = (r_i, we_i, rd_i, p_i)$

r_i : released time for the transaction, which represent the arrival time.

we_i : the estimated worst case execution time.

rd_i : the relative deadline, it indicates the requirement to complete the transaction before the instant deadline.

p_i : the priority, of transaction, which depend on the transaction relative deadline.

VII. ADRTDBS SYSTEM UNITS

An ADRTDBS system capable of executing transactions with timely constraint in distributed environment. The system consists of ten of working units, and each server contains copy of program for these units. Figure (3) illustrate the prototype of the ADRTDBS system.

• Transaction Admission Control (TAC) Unit:

This unit receives transaction request from database servers and clients. This unit provides a database interface to the application. This interface consists of a data manipulated language, in which the user (application) can query and manipulate data elements.

• Index Management (IM) Unit:-

It is used to maintain an index for all tuples in the database. It is capable of transforming a database key into the memory address of the tuple correspondent to the database key.

• Memory Management (MM) Unit

This unit is responsible for memory allocation of tuples and database indexes.

• Transaction Management (TM) Unit :-

This unit responsible of managing transactions coming from admission control unit TAC and transmit it to scheduler unit TS to schedule them according to the proposed algorithm. This unit provides required data to each transaction and controls and assembles results for each request. This unit also controls and manages other units and calculate deadline for each transaction.

A deadline function computes the execution time for each transaction and predicts the deadline according to the following equation:

$$TD = RL(T) + Pr_Ex(T) * SF$$

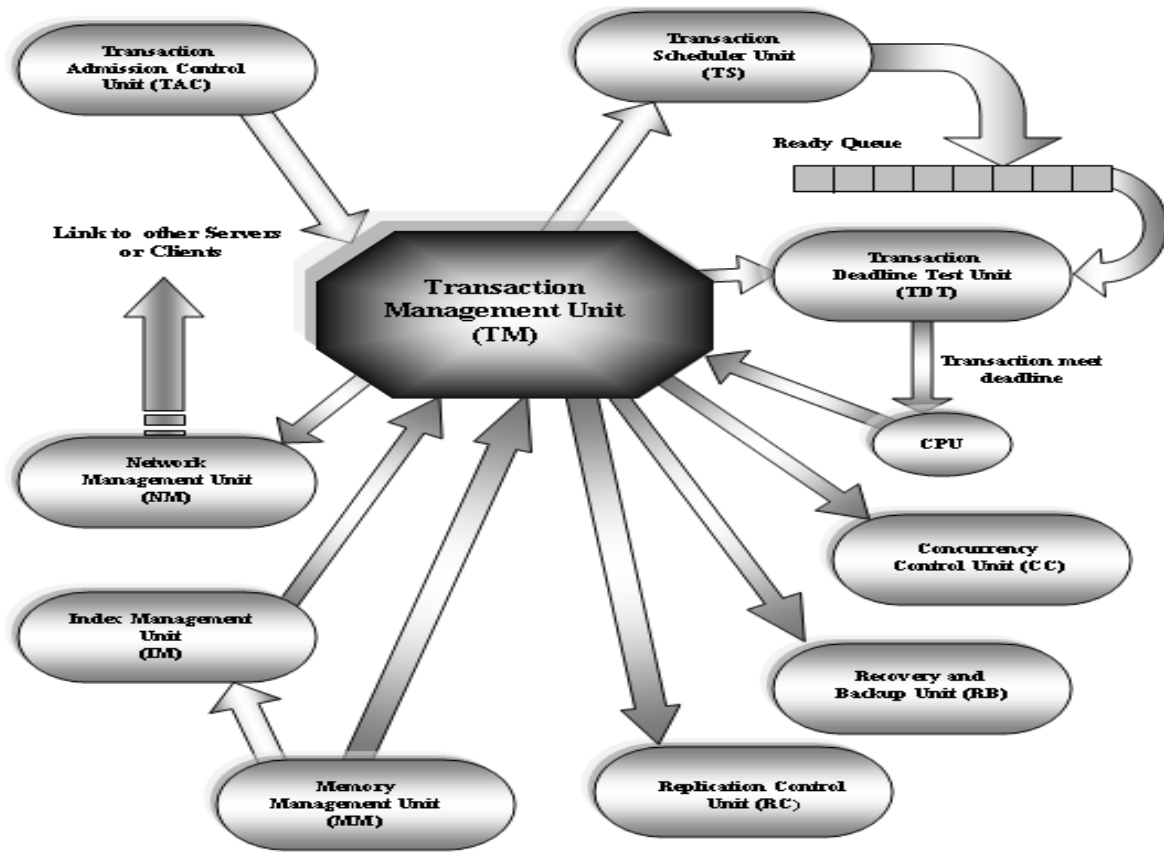


Figure (3) Units of Proposed ADRTDBS System

Where

$RL(T)$: release time for transaction T

$Pr_Ex(T)$: Predict execution time for T and this time can be computed by

$$Pr_Ex(T) = (T_operation + T_update) * N_op + T_cc$$

Where

$T_operation$: time to process an operation

T_update : time to update data

N_op : number of operation

T_cc : communication cost

SF : slack factor $15 \geq SF \geq 1$

• Transaction Scheduler (TS) Unit

This is responsible for scheduling transactions. This unit maintains the list of transactions in a ready queue and releases the next transaction when the previous is completed and give it to the CPU. The ready queue is organized according to the transaction priority. Each transaction is characterized by a

deadline which defines its urgency with respect to the other transactions of real time application. The higher priority is given to transaction with minimum deadline according to the scheduling algorithm EDTDF (Earliest Data or Transaction Deadline First). If the system cannot complete transaction before its deadline, the transaction is aborted.

The algorithm is work like this:

1. Receive transaction from TM unit.
2. Determine if the transaction contains temporal data.
3. Define the deadline for the temporal data (DD).
4. Compute the deadline for transaction (TD).
5. Compute the Final Transaction Deadline (FTD) by
if the transaction contain temporal data then
 $FTD = \text{Min}(TD, DD)$
else $FTD = TD$
6. Put the transaction in the ready queue with transaction with earliest deadline at the head of the queue.

• Transaction Deadline Test (TDT) Unit

This unit is responsible of decision that a transaction is aborted whenever it is found to have missed its deadline. So, for each transaction periodically checks whether or not the transaction will be able to meet its deadline taking into consideration the fact that the transaction has to update the data object in its write-set at each database. If the system's current time plus the time to update all data objects in a transaction's write-set is greater than the transaction's deadline, it means that this transaction will not be able to commit before its deadline is reached. In order not to be waste any system resources, the transaction will be aborted and removed from the system. Figure (4) illustrate the work of this unit.

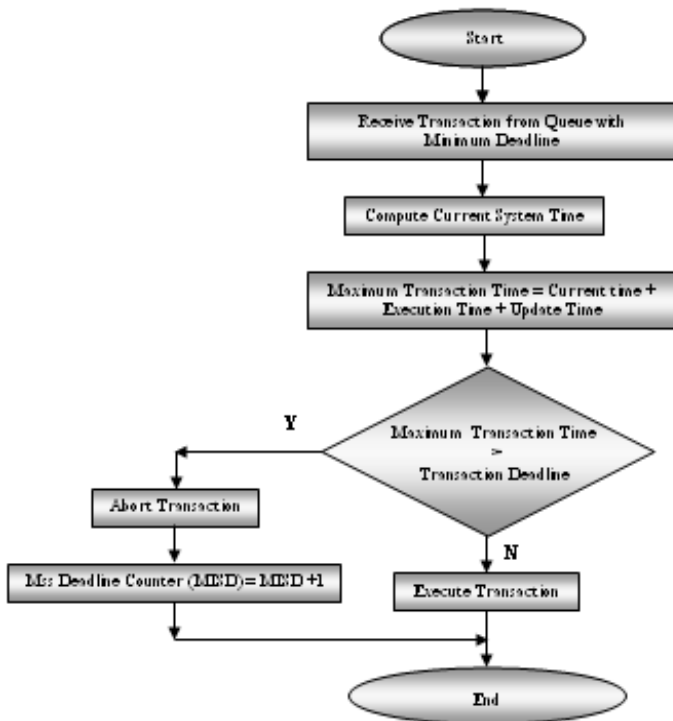


Figure (4) Transaction Deadline Test unit

• Concurrency Control (CC) Unit

This unit is responsible of synchronous execution for more than one transaction that require execution on same database at same time. In this work a 2PL (Two Phase Locking) protocol was used to control concurrency. 2PL protocol follows the following three rules:

- When the protocol receive a lock request, it tests whether the requested lock conflicts with another lock that is already set. If so, it queues the lock request. If not, it responds to the lock request by setting the lock.

- The 2PL protocol locks a data item only once, it cannot release the lock until the DM (Data Manager) has completed processing of the lock's corresponding operation.
- The 2PL protocol releases a lock for a transaction, it may not subsequently allow any lock for the same transaction.

• Network Management (NM) Unit

This unit is responsible of managing the transfer of data between servers Depending on the TCP/IP Protocol which consider the best protocol that provide high speed for sending and receiving data over the network.

Database server communicates among each other via 1-to-n communication which is consider as group communication. Reliable broadcast parameter of this communication ensures that a message sent by a correct database server, or delivered by a correct database server, is eventually delivered to all correct database servers.

• Replication Control (RC) Unit

This unit controls all updating on the local database. This unit broad changes on database to remain copies of database on servers connected by network in synchronizing manner by using 2PC (Two Phase Committing).

Replication requires to have a specific site – the main copy – associated with each data item. The clients must send their requests to one particular server. This server is the main copy. Because there is only one server executing the transactions, there are no conflicts across the servers. Any update to the data item must be first sent to the main copy where it is processed. The main copy then propagates the update (or its results) to all other sites. This approach is used in ADRTDBS system to minimize conflicts among transactions executed over replicated data.

The steps for this technique are the following:

1. The transaction starts at the primary copy site.
2. Read operations are executed locally.
3. The result of write operations are broadcast to the other sites. (backups). (i.e. update every where).
4. The main copy site starts the Two Phase Commitment Protocol (2PC).
5. The transaction is committed on all sites.

• Recovery and Backup (RB) Unit

This unit makes backups of database and recovering it when required. This unit maintains information on database when an error occur on computer or database or when we need copy the database on more than one computer. The famous manner of back up and recovery is export and import.

The steps of export and import are illustrated in figure (5).

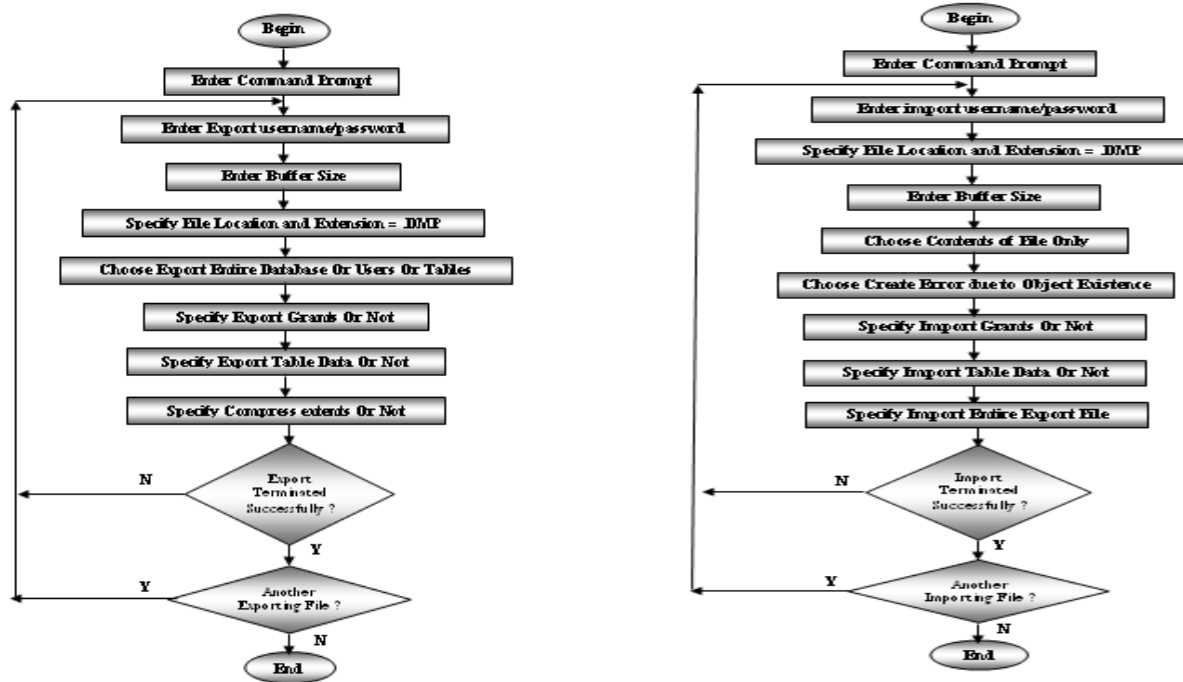


Figure (5) : a- Export Data algorithm, b- Import Data algorithm

VIII. SYSTEM TEST

The system has been tested on Alrafidain bank. Many transactions were take into consideration, (add customer , close customer account, draw money, transfer funds, currency change, etc).

. Transfer Fund Transaction: If we transfer fund between customers the system demand many steps illustrated in the following figure (6). The figure also shows that every step (or a number of steps) of the unit within a system ADRTDBS. The time parameters for this transaction are illustrated in table (1).

Table (1) Deadline Computation of Transfer Fund Transaction

No of Tables That Effected	6
No of Record That Effected	6
No of Fields That Effected	64
Type of Operation	Updating & Adding
T_operation	0.2778 mill sec.
T_update	0.1388 mill sec.
T_cc	2 mill sec.
N_op	6 mill sec.
Pr_Ex(T) = (T_operation + T_update) * N_op+ T_cc	4.5 mill sec.
RL(T)	0 mill sec.
SF	2.8
TD = RL(T) + Pr_Ex(T) * SF	12.6 ~ 12 mill sec.

IX. PERFORMANCE EVALUATIONS

In conventional distributed database systems, performance is primarily measured by the number of transactions completed within a unit time. In distributed real time database systems, timing and criticality characteristics of transactions must be taken into account. So performance depends on many other criteria, which are related to real time. Some of these criteria are the number of transactions that missed their deadline, average tardy time, etc. In this work, the performance metric employed is the percentage of transactions that missed their deadlines (%miss) in the total number of transactions that were submitted to ARTDDBS system during the session period :

$$\text{Miss ratio} = \frac{\text{No. of missed deadline transactions}}{\text{Total transactions}} * 100$$

Also we measure the total CPU utilization. And develop performance measurement take into consideration database reside in main memory consisting of 72 organism data. (Table 2) shows the model parameters and their baseline values.

We take a sample of 70 transactions of this application distributed as (add new customer, updating customer information, close account, deposit money, query about account, transfer fund, display customer information). The time of execute this transactions are 293 millisecond, and the transactions missed their deadline 14.285%.

$$\text{Miss Ratio} = 10 / 70 * 100 = 14.285$$

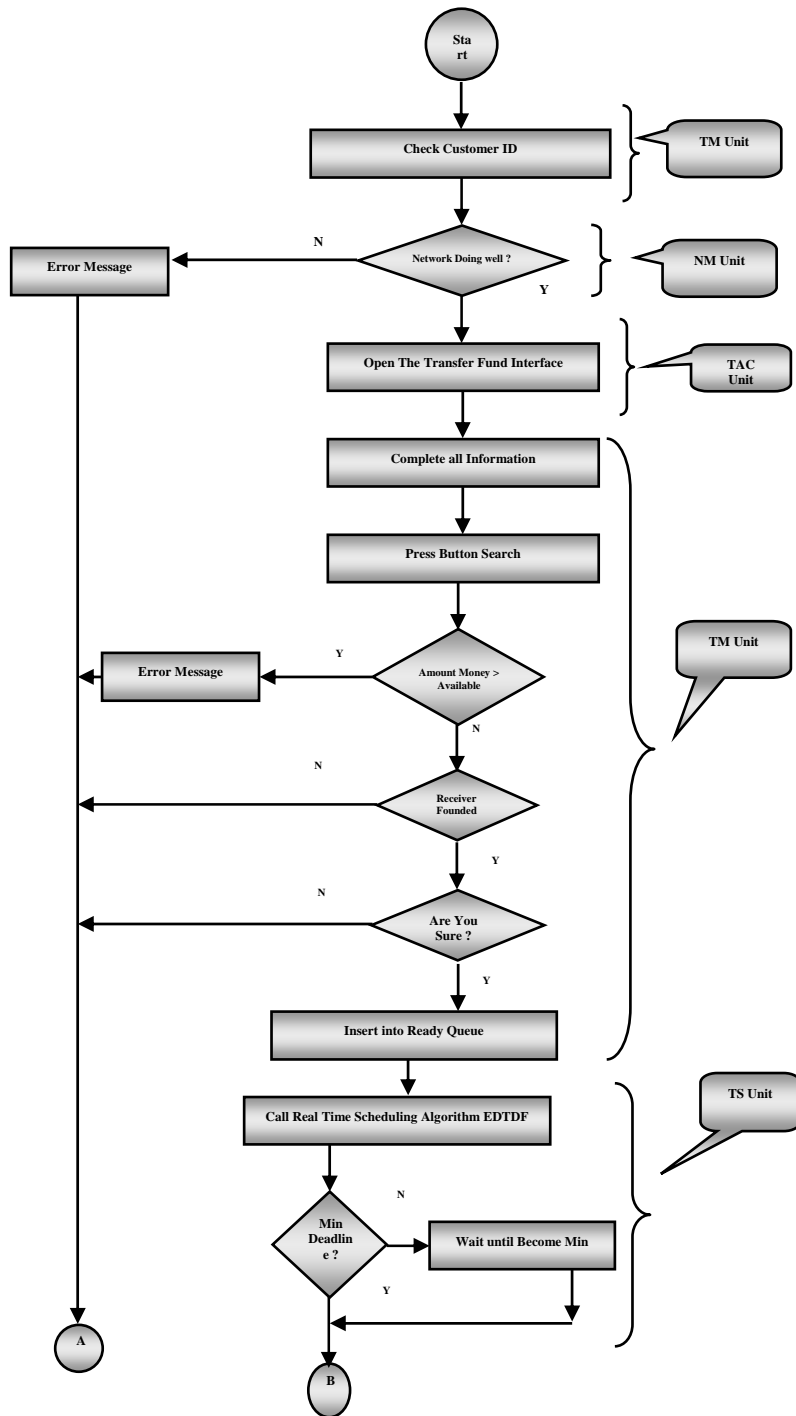


Figure (6) : Transfer Fund Execution in ADRTDBS

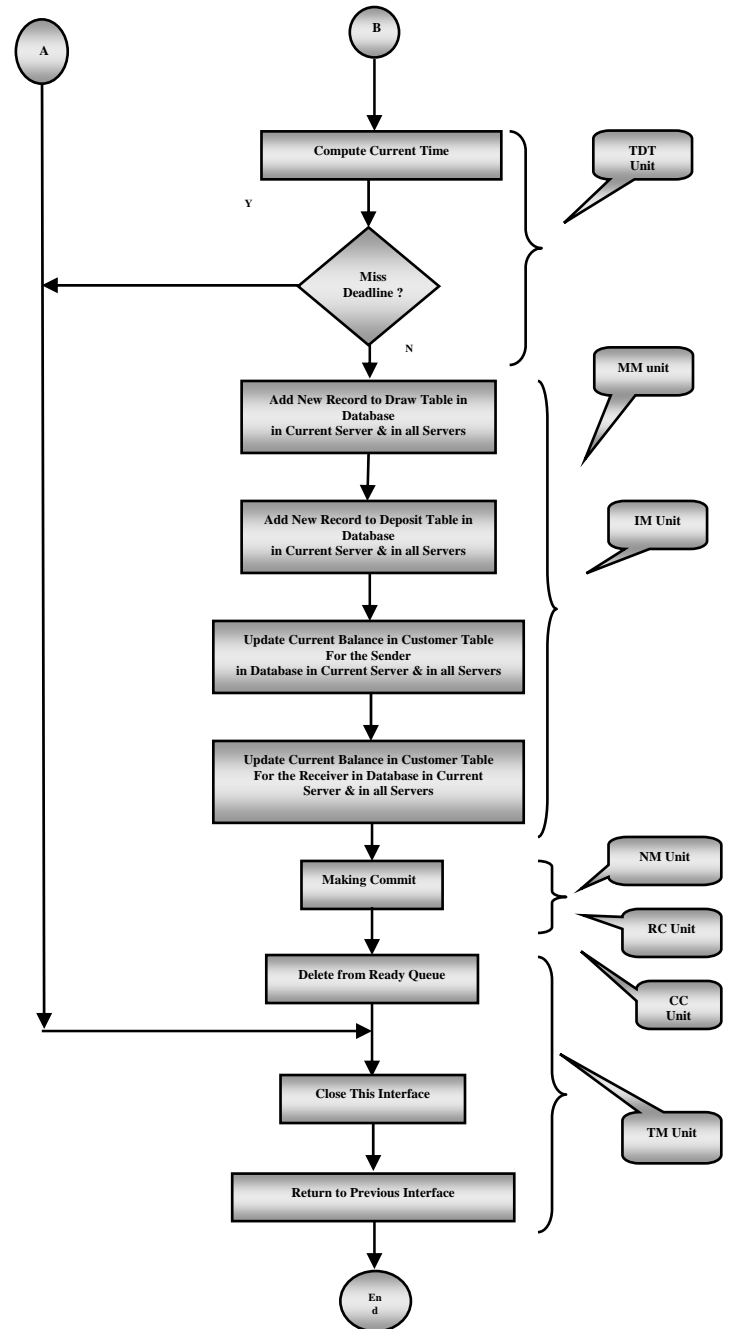


Figure (6) : Cont. Transfer Fund Execution in ADRTDBS

Table (2) Performance Evaluation

Parameter	Base line value
System	
Number of servers	2
Number of clients for each server	2
Communication cost	2 mill sec.
Database	
Number of local databases in each site	1 Database
Number of objects in local database in each site	11 tables, and 5 views
Database size	16 data object per local database
Concurrency control	2PL (two phase locking)
Fraction of temporal data object	0.1
Transaction	
Transaction size	3 to 5 operation uniform distributed
Proportion of write operations	0.35
Slack Factor Range	1...15 the slack factor is uniformly distributed in the slack range (we use 2.8)
CPU	
CPU scheduling	EDTDF (earliest data or transaction deadline first)
CPU time to process an operation	1...6 mill sec.
CPU time to commit changes on database	1 mill sec.
CPU time to rollback changes on database	1 mill sec.
CPU time to update a data object	6 mill sec.

The figure (7) illustrates measuring of Miss Ratio of this system and figure (8) illustrates the CPU utilization.

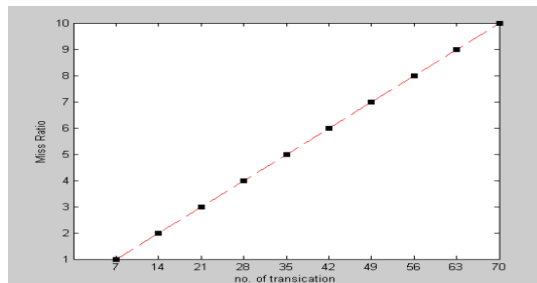


Figure: (7) Miss Ratio of this system.

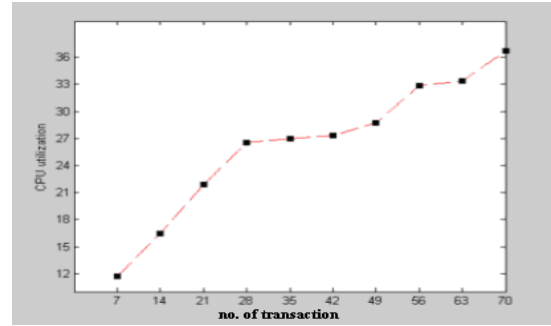


Figure (8) CPU utilization

CONCLUSIONS AND FURTHER WORKS

In this paper a model of proposed distributed real-time database system was designed, as this system has the ability to execute real time transactions in distributed environment. The proposed system uses full replication method, replication of the same database on all sites.

The replication of the whole database at every site in this proposed system improves availability remarkably because the system can continue to operate as long as at least one site is functioning well. It also improves performance of retrieval for global queries, because the result of such a query can be obtained locally from any site, hence a retrieval query can be processed at the local site where it is submitted, if that site includes a server module. Replication of data from the remote site to the local site makes the data available on the local site and minimizes the response execution time which very suitable to distributed real time databases environment. Also, by maintaining multiple copies, it becomes possible to provide better protection against corrupted data.

using 2PL (Two Phase Locking) protocol help to control synchronize execution of transactions in this proposed system in two cases:

- When there is a request to execute same transactions from some clients on one server at same time.
- When there is a request to execute same transactions from some servers on one server at same time.

As a further work we suggest to use new encryption algorithms to increase the security of system. Use another distributed manner of database like hybrid and then see how it is compatible with distributed real time database.

REFERENCES

- [1] Al-Kinany, E. A., "Heterogeneous System Design for Distributed Database Systems", Msc., Alrasheed College, University of Technology, 2005.
- [2] Bouzeffrane, S. S.; Kaiser, C., "Distributed Overload Control for Real Time Replicated Database Systems", Cedric

- Laboratory, Conservatoire National des arts et Metiers, FRANCE, 2002.
- [3] Brohede, M.; Andler, S. F., "Using Distributed Active Real Time Database Functionality in Information Fusion Infrastructures", University of Skode, SE-54128, Sweden. <http://cgi.omg.org/docs/ptc/01-08-34.pdf>, 2005.
 - [4] Coulouris, G.; Dollimore, J.; Kindberg, T., "Distributed Systems Concepts and Design", Fourth Edition, ADDISON-WESLEY PUBLISHING, 2005.
 - [5] Deborah, S.; Henry, R., "Oracle9i Net Services Administrator's Guide", Release 2 (9.2), Oracle Corporation, 2002.
 - [6] Deng, G.; Schmidt, D. C., Nechypurenko A.; Gokhale, A., "Developing Distributed Real Time and Embedded Systems with Modeling Tools and component Middleware: A Case Study", Department of EECS, Vanderbilt University, Siemens Corporate Technology, Germany, 2005.
 - [7] Graham, M. H., "Issues in Real Time Data management", Technical Report, CMU/SEI-91-TR-017, ESD-91-TR-017, 1991.
 - [8] Lindstrom, J.; Niklander, T.; Raatikainen, K., "A Distributed Real Time Main Memory Database Telecommunication", Department Computer Science, University of Helsinki. www.imtrg.me.metu.tr/publications/paper.pdf, 1998.
 - [9] Peddi, P.; Dipippo, L. C., "A Replication Strategy for Distributed Real Time Object Oriented Database", Proceedings of the Fifth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing, Washington, D.C., 2002.
 - [10] Ramamritham, K., "Real Time databases", International Journal of Distributed and Parallel Databases 1(2), 199–226, 1993.
 - [11] Ravindran, B.; Anderson, J. S.; Jensen, E. D., "On Distributed Real Time Scheduling in Network Embedded Systems in the Presence of Crash Failures", ECE Dept. Virginia Tech, VA 24061, USA, The MITRE Corporation, Bedford, MA 01730, USA, 2006.
 - [12] Shanker, U., "Some performance issues in distributed real time database systems", M.sc., Department of Computer Science, University of Virginia, TesiOnline. USA, 2000.
 - [13] Soparker, N.; Lery E.; Korth, H. F.; Silberschatz, A., "Adaptive Commitment for Distributed Real Time Transactions", Work Partially Supported By NSF Grant IRI-8805215, and by a Grant from IBM Corporation, USA, 1994.
 - [14] Syberfeldt, S., "Optimistic Replication with forward Conflict Resolution in Distributed Real Time Database", Ph.D., Department of Computer and Information Science, Linköpings University, Sweden. Printed by LiU-tryck, Linköping, 2007.
 - [15] Tao, J.; Williams, J. G., "Concurrency Control and Data Replication Strategies for Large Scale and Wide Distributed Database", University of Pittsburgh, IEEE, 0-7695-0996-7/01, 10.00, 2001.
 - [16] Wei, Y.; Prasad, V.; Son, S. H., "Qos Management of Real Time Data Stream Queries in Distributed Environments", Department of Computer Science, University of Virginia. www.cs.virginia.edu/~radb/sch.html, 2006.
 - [17] Wei, Y.; Son, S. H.; Stankovic, J. A., "Maintaining Data Freshness in Distributed Real Time Databases", USA, This Work was Supported in Part by NSF Grant IIS-0208758, CCR-032,60, and CCR 0098269, 2003.
 - [18] Wolfe, V. F.; Black, J.; Thuraisingham, B.; Krupp, P., "Real Time Method Invocation in Distributed Environment", USA, The MITRE Corporation Before MA, USA, This Work is Partially supported by the US National Science Foundation, US NAVAL Undersea Warfare Center, 1994.
 - [19] Woonchul K., Sang H., and John A., "DRACON: QoS Management for Large-Scale Distributed Real-Time Databases", JOURNAL OF SOFTWARE, VOL. 4, NO. 7, 2009.
 - [20] Zharkov, A., "Performance Evaluation of Transaction Handling Policies on Real Time Database Management System Prototype", Proceeding of the Spring Young Researcher's Colloquium On Database and Information Systems SYRCODIS, MOSCOW, Russia, 2007.
 - [21] Zharkov, A., "On Using Materialized Views for Query Execution in Distributed Real Time Database Management System", Proceedings of the Spring Young Researcher's Colloquium On Database and Information Systems SYRCODIS, St.-Petersburg, Russia, 2008.



Dr. Dhuha Basheer Abdullah Albazaz /Asst. Prof / computers Sciences Dept. / College of Computers and Mathematics / University of Mosul. She has a Ph.D. degree in Computer Sciences since 2004. Specific Specialist in Computer Architecture and Operating System. Supervised many Master degree students in operating system, computer architecture, dataflow machines, mobile computing, real time, distributed databases. She has three Phd. Students in FPGA field, distributed real time systems, and Linux clustering. She also leads and teaches modules at both BSc, MSc, and Phd. levels in computer science. Also she teaches many subjects for Ph.D. and master students.

Person Identification System using Static-dynamic Signatures Fusion

Dr. S.A Daramola

Department of Electrical and Information Engineering
Covenant University
Ota, Nigeria

Prof. T.S Ibiyemi

Department of Electrical Engineering
University of Ilorin
Ilorin, Nigeria

Abstract—Off-line signature verification systems rely on static image of signature for person identification. Imposter can easily imitate the static image of signature of the genuine user due to lack of dynamic features. This paper proposes person identity verification system using fused static-dynamic signature features. Computational efficient technique is developed to extract and fuse static and dynamic features extracted from offline and online signatures of the same person. The training stage used the fused features to generate couple reference data and classification stage compared the couple test signatures with the reference data based on the set threshold values. The system performance is encouraging against imposter attacker in comparison with previous single sensor offline signature identification systems.

Keywords- fused static-dynamic signature; feature extraction; forgeries

I. INTRODUCTION

Person identity verification is a problem of authenticating individual using physiological or behavioral characteristics like face, iris, fingerprint, signature, speech and gait. Person identification problem can be solved manually or automatically. Biometric systems automatically use biometric trait generated from one or two sensors to validate the authenticity of a person. Automatic person identity verification based on handwritten signature can be classified into two categories: on-line and off-line, differentiated by the way signature data is acquired from the input sensor. In off-line technique, signature is obtained on a piece of paper and later scanned to a computer system while in on-line technique, signature is obtained on a digitizer thus making dynamic information like speed, pressure available while in offline only the shape of the signature image is available [1] [2].

In this paper, combination of offline and online signatures are used for person identification. The process involves verification of a signature signed on both paper and electronic digitizer concurrently. Therefore the physical present of the signer is required during the period of registration and verification. This type of system is useful particular in the bank while the physical present of the holders of saving current are required before money can be withdrawn. In Nigeria many banks manually identify holder of saving current using face and static signature, in the process genuine users

are rejected as imposters because of high intra-class variation in signatures. Frauds as result of signature forgeries must be prevented particular among closely resemble people. Fusion of dynamic and static signature will strengthen the identification of people physically in paper documentation environment.

A detailed review of on-line signature verification including summary of off-line work until the mid 1990's was reported in [1] [2]. Alessandro et al. [3] proposed a hybrid on/off line handwritten signature verification system. The system is divided into two modules. The acquisition and training module use online-offline signatures, while the verification module deals with offline signatures. Soltane et al [4] presented a soft decision level fusion approach for a combined behavioral speech-signature biometrics verification system. And Rubesh et al [5] presented online multi-parameter 3D signature and cryptographic algorithm for person identification. Ross et al [6] presented hand-face multimodal fusion for biometric person authentication while Kiskus et al [7] fused biometrics data using classifiers.

From the approaches mentioned above, some of the authors used online signature data to strengthen the system performance either at registration or training stage, while others combined online signature data with other biometric modalities data like speech, face as means of person identification. The system proposes in this novel frame work is based on fusion of static and dynamic signature data at feature level for person identification. The universal acceptance of signature, compatibility of offline and online signature features make the proposed system more robust, accurate and friendly in comparison with other previous multi biometric modalities systems or single sensor offline system for person identification.

Section 2 provides the description of the system, the signature preprocessing and feature extraction and fusion technique. Also in section 2, the signature training, threshold selection and classification are presented. Section 3 shows the experimental results and finally, conclusions are drawn in section 4.

II. PROPOSED SYSTEM

The system block diagram is shown in Fig. 1. The offline and online data are collected at the same time from the same user during registration/training and verification exercises. Also the offline signature are preprocessed to remove unwanted noise introduced during scanning process whereas, the online signatures are not preprocess in order to preserve the timing characteristics of the signature. Discriminative static and dynamic features are extracted separately from offline and online signature respectively. At the feature level the two signatures are fused together to obtain a robust static-dynamic features. These features are used to generate couple reference data during training and for signatures classification.

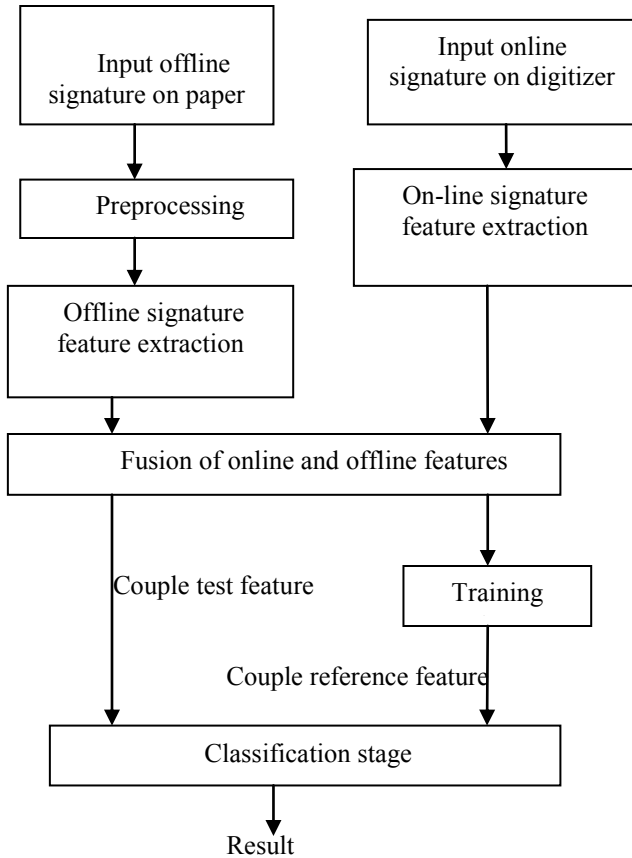


Figure1: System block diagram

A. Data Acquisition

The signature database consists of a total number of 300 offline and 300 online handwritten genuine signature images and 100 forged signatures. The genuine signatures are collected from 50 people. Each of the users contributed 6 offline and 6 online signature samples. The 100 skilled forgeries consist offline and online signatures, they are collected from 25 forgers and each of the forgers contributed 4 samples. The raw signature data available from our digitizer consists of three dimensional series data as represented by (1).

$$S(t) = [x(t), y(t), p(t)]^T \quad t = 0, 1, 2, \dots, n \quad (1)$$

where $(x(t), y(t))$ is the pen position at time t , and $p(t) \in \{0, 1, \dots, 1024\}$ represents the pen pressure.

B. Offline Signature Preprocessing

The scanned offline signature images may contain noise caused by document scanning and it has to be removed to avoid errors in further processing steps. The gray-level image is convolved with a Gaussian smoothing filter to obtain a smoothed image. The smoothed gray image is converted into binary image and then thinned to one pixel wide.

C. Offline Feature Extraction

The feature extraction algorithm for the static signature is stated as follows:

- (1) Locate signature image bounding box.
 - (i) Scan the binary image from top to bottom to obtain the signature image height.
 - (ii) Scan the binary image from left to right to obtain the signature image width.
- (2) Centralization of the signature image.
 - (i) Calculate centre of gravity of the signature image using (2).

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x(i),$$

$$\bar{y} = \frac{1}{N} \sum_{j=1}^N y(j).$$

(2)

- (ii) Then move the signature image centre to coincide with centre of the predefined image space.
- (3) The image is partitioned into four sub-image parts.
 - (i) Through point \bar{x} make a horizontal splitting across the signature image.
 - (ii) Through point \bar{y} make a vertical splitting across the signature image.
- (4) Partition each of the sub-image parts into four rectangular parts.

- (i) Locate the centre of each of the sub-image parts using (2).
- (ii) Repeat step 3(i) and 3(ii) for each of the sub-image parts in order to obtain a set of 16 sub-image parts.
- (5) Partition each of the 16 sub-image parts into four signature cells
 - (i) Locate the centre of each of the sub-image parts using (2).
 - (ii) Repeat step 3(i) and 3(ii) for each of the sub-image parts in order to obtain a set of 64 sub-image cells.
- (6) Calculate the angle of inclination of each sub-image centre in each cell to the lower right corner of the cell.
 - (i) Locate the centre of each of the 64 sub-image cells using (2)
 - (ii) Calculate the angle that each centre point makes with the lower right corner of the cell.

The feature extracted at this stage constitutes the offline feature vector, which is represented as: $F = f_1, f_2, f_3, f_4, \dots, f_{64}$. The details and diagrams of the feature extraction are given in [8].

D. Online Signature Extraction

Three on-line signature features are extracted at each sampling point from the raw data. The features are $\Delta p/\Delta x$, $\Delta p/\Delta y$ and v . Δx corresponds to change of x between two successive sampling points, Δy corresponds to change of y between two successive sampling points, Δp corresponds to change of p between two successive sampling points, $\Delta p/\Delta y$ corresponds to ratio of Δp to Δy , $\Delta p/\Delta x$ corresponds to ratio of Δp to Δx and v corresponds to change of speed between two successive sampling points [9]. These features are obtained using (3), (4) and (5).

$$v = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (3)$$

$$\frac{\Delta p}{\Delta y} = \frac{p(t) - p(t-1)}{y(t) - y(t-1)} \quad (4)$$

$$\frac{\Delta p}{\Delta x} = \frac{p(t) - p(t-1)}{x(t) - x(t-1)} \quad (5)$$

E. Fusion of Offline and Online features

This technique is designed to compute fused feature vector, which contains information from both the offline and online signatures and used this feature vector for subsequent processing. Information from two input sensors can be combined at data acquisition stage, feature extraction stage or at decision level stage. The accuracy of the system also depends on the level of fusion and the discriminative ability of fused data. In [4] the fusion of voice and signature data was

done at decision level. While in [6] fusion of hand and face was done at feature level. In this work offline feature is combined with online feature at feature level. The compatibility of the signature data from the same person, from different sensors made the fusion possible without any loss of information. The steps involve are stated as follows: given that extracted static feature is $F = f_1, f_2, f_3, f_4, \dots, f_{64}$. The mean and variance of the feature vector are calculated using (6) and (7) respectively

$$\overline{\mu_{off}} = \frac{1}{N} \sum_{i=1}^N f_i \quad (6)$$

$$\sigma_{off}^2 = \frac{1}{N} \sum_{i=1}^N (f_i - \overline{\mu_{off}})^2 \quad (7)$$

The fused features are obtained by normalized each of the extracted online features $(v, \frac{\Delta p}{\Delta y}, \frac{\Delta p}{\Delta x})$ using the variance of the offline feature (σ_{off}^2) . The three fused features (SF1, SF2 and SF3) become: $\frac{v}{\sigma_{off}^2}, \frac{\Delta p}{\Delta y \sigma_{off}^2}, \frac{\Delta p}{\Delta x \sigma_{off}^2}$.

F. Training and Threshold Setting

Each of the registered users submitted 12 genuine signatures to the system, out of which 8 signatures are fused together to generate 4 couple reference features. These features are used to generate 6 distance values by cross-aligning the couple reference features to the same length using Dynamic Time Warping (DTW). These distance values are used to measure the variation within each of the user's signatures, so as to set user-specific threshold for accepting or rejecting a couple test signatures. Given four couple reference signature samples R1, R2, R3 and R4, these features are cross aligned to obtain 6 distance values as shown in Fig.2. The mean (m_k) and standard deviation (σ_k) of the distances: d_{12} , d_{13} , d_{14} , d_{23} , d_{24} and d_{34} are calculated and used to set the threshold (t_k) for each of the users based on each of the fused features as given in (8).

$$0 \geq t_k \leq m_k + \sigma_k \quad (8)$$

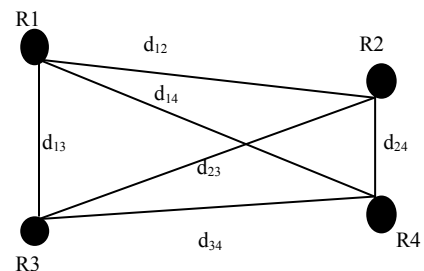


Figure2. Cross-alignment of Couple reference features

G. Classification of Couple Signature Features

Whenever a couple test signatures (offline and online) come into the system, the fused feature vector of the couple test signatures is pair-wise aligned with each of the four couple reference features using DTW. Four distance values are obtained as shown in Fig.3. The distance (d_i) of the couple test feature (FT) from the four couple reference features R1, R2, R3 and R4 is calculated using (9).

$$dt = \frac{f_{T1} + f_{T2} + f_{T3} + f_{T4}}{4} \quad (9)$$

If d_i is within the assigned threshold value then the fused test signature is assigned a pass mark otherwise it has no pass mark. Finally a decision by the system in accepting or rejecting a couple test signatures is based on total pass mark it obtained based on the three fused features.

III. EXPERIMENTAL RESULTS

Experiments have been conducted to evaluate the discriminative ability of each of the fused features against forgers attack. Also the proposed system is tested based on the three new fused features. Total number of 150 fused signatures made up of 100 genuine signature features and 50 skilled forgery features are collected from 75 people are tested. The performance evaluation is based on False Acceptance Rate (FAR) and False Rejection Rate (FRR). Table 1 shows the results of the performance of these fused features in comparison with previous single offline features. Table 2 shows the proposed system FAR for skilled forgeries and the FRR for genuine signatures.

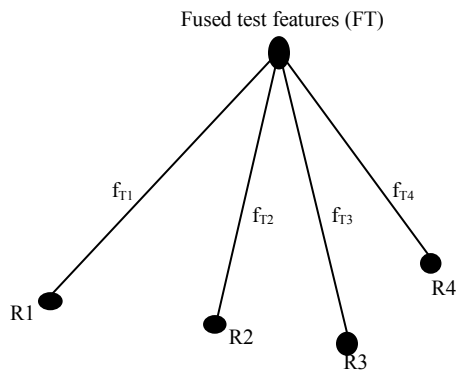


Figure 3. Distance between couple reference features and couple test feature

TABLE1: OFFLINE FEATURES IN COMPARISON WITH THE PROPOSED FUSED FEATURES.

Type	Feature	FRR	FAR
Some previous related offline features [8][10]	Pixels normalized angle relative to the cell lower right corner	1.250	2.500
	Image centre angle relative to the cell lower right corner	2.500	2.500
	Vertical centre points	7.500	8.750
	Horizontal centre points	6.250	7.500
Proposed fused offline-online features	SF1	0.150	0.120
	SF2	0.120	0.052
	SF3	0.100	0.080

TABLE2: FAR AND FRR RESULTS OF THE PROPOSED SYSTEM

Type	Total	Accepted	Rejected	FAR	FRR
Individual genuine fused features	100	95	5	—	0.05
Individual skilled fused forgeries	50	1	49	0.02	—

IV. CONCLUSION

This paper has proposed a person identification system using fused signature features from two biometric sensors. Fused signature feature is used to strengthen verification system in paper documentation environment like banks where the present of the account holders are required for transaction. Signature is universally accepted, this make the proposed system more friendly and acceptable in comparison with

others biometric traits combination. The experimental results have shown that fused signature identification method is more accurate in comparison with previous single sensor offline signature identification techniques.

REFERENCES

- [1] R. Plamondon and S.N. Srihari "On-line and off-line handwriting recognition: a comprehensive Survey", IEE trans. on Pattern Analysis and Machine Intelligence, vol. 22, No.1, pp. 63-84, 2000.
- [2] F. Leclerc and R. Plamondon, "Automatic verification and writer identification: the state of the art 1989-1993", International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, pp. 643 – 660, 1994.
- [3] A. Zimmer and L.L. Ling "A hybrid on/off line handwritten signature verification system", Proc. of the seventh International Conference on Document Analysis and Recognition, 2003.
- [4] S. Mohamed, G. Nouredine and D. Nouredine "Soft decision level fusion approach to a combined behavioral speech-signature biometrics verification", International journal of Biometrics & Bioinformatics, vol. 4, issue 1. 2009.
- [5] P.M. Rubesh, G. Bajpai and V. Bhaskar "Online multi-parameter 3D signature verification through Curve fitting", International Journal of Computer Science and Network Security, vol.9, No.5, pp 38-44. 2009.
- [6] A. Ross and R. Govindarajan "Feature level using hand and face biometrics", Proc. of SPIE Conference on Biometric Technology for Human Identification, vol. 5779, pp.196-204, 2005.
- [7] D. R. Kisku, P. Gupta and J. K. Sing "Offline signature identification by fusion of multiple classifiers using statistical learning theory", International Journal of Security and Its Applications, vol.4, No.3, 2010.
- [8] S. Daramola and S. Ibiyemi "Novel feature extraction technique for offline signature verification", International Journal of Engineering Science and Technology, vol (2)7, pp 3137-3143, 2010.
- [9] S. A. Daramola and T.S Ibiyemi "Efficient On-line Signature verification" International Journal of Engineering & Technology, vol 10, No. 4. pp 48-52. 2010.
- [10] M. Banshider, R .Y Santhosh and B .D Prasanna "Novel features for off-line signature verification" International Journal of Computers, Communications & Control, vol. 1 , No. 1, pp. 17-24. 2006.

AUTHORS PROFILE

Dr. S.Adebayo Daramola obtained Bachelor of Engineering from University of Ado-Ekiti, Nigeria, Master of Engineering from University of Port Harcourt, Nigeria and PhD from Covenant University, Ota, Nigeria. His research interests include Image processing and Cryptography.

Prof. T.S Ibiyemi is a Professor in Computer Engineering. He has more than 30 years teaching and research experience; he has many papers in local and international journals. His research interests include Image processing, Multimedia and Processors architecture.

Short term flood forecasting using RBF static neural network modeling a comparative study

Rahul P. Deshmukh

Indian Institute of Technology, Bombay
Powai, Mumbai
India.

A. A. Ghatol

Former Vice-Chancellor
Dr. Babasaheb Ambedkar Technological University,
Lonere, Raigad, India.

Abstract—The artificial neural networks (ANNs) have been applied to various hydrologic problems recently. This research demonstrates static neural approach by applying Radial basis function neural network to rainfall-runoff modeling for the upper area of Wardha River in India. The model is developed by processing online data over time using static modeling. Methodologies and techniques by applying different learning rule and activation function are presented in this paper and a comparison for the short term runoff prediction results between them is also conducted. The prediction results of the Radial basis function neural network with Levenberg Marquardt learning rule and Tanh activation function indicate a satisfactory performance in the three hours ahead of time prediction. The conclusions also indicate that Radial basis function neural network with Levenberg Marquardt learning rule and Tanh activation function is more versatile than other combinations for RBF neural network and can be considered as an alternate and practical tool for predicting short term flood flow.

Keywords—component; Artificial neural network; Forecasting; Rainfall; Runoff;

I. INTRODUCTION

The main focus of this research is development of Artificial Neural Network (ANN) models for short term flood forecasting, determining the characteristics of different neural network models. Comparisons are made between the performances of different parameters for Radial basis function artificial neural network models.

The field engineers face the danger of very heavy flow of water through the gates to control the reservoir level by proper operation of gates to achieve the amount of water flowing over the spillway. This can be limited to maximum allowable flood and control flood downstream restricting river channel capacity so as to have safe florid levels in the river within the city limits on the downstream.

By keeping the water level in the dam at the optimum level in the monsoon the post monsoon replenishment can be conveniently stored between the full reservoir level and the permissible maximum water level. Flood estimation is very

essential and plays a vital role in planning for flood regulation and protection measures.

The total runoff from catchment area depends upon various unknown parameters like Rainfall intensity, Duration of rainfall, Frequency of intense rainfall, Evaporation, Interception, Infiltration, Surface storage, Surface detention, Channel detention, Geological characteristics of drainage basin, Meteorological characteristics of basin, Geographical features of basin etc. Thus it is very difficult to predict runoff at the dam due to the nonlinear and unknown parameters.

In this context, the power of ANNs arises from the capability for constructing complicated indicators (non-linear models). Among several artificial intelligence methods artificial neural networks (ANN) holds a vital role and even ASCE Task Committee Reports have accepted ANNs as an efficient forecasting and modeling tool of complex hydrologic systems[22].

Neural networks are widely regarded as a potentially effective approach for handling large amounts of dynamic, non-linear and noisy data, especially in situations where the underlying physical relationships are not fully understood. Neural networks are also particularly well suited to modeling systems on a real-time basis, and this could greatly benefit operational flood forecasting systems which aim to predict the flood hydrograph for purposes of flood warning and control[16].

A subset of historical rainfall data from the Wardha River catchment in India was used to build neural network models for real time prediction. Telematic automatic rain gauging stations are deployed at eight identified strategic locations which transmit the real time rainfall data on hourly basis. At the dam site the ANN model is developed to predict the runoff three hours ahead of time.

In this paper, we demonstrate the use of Radial basis function neural network (RBF) model for real time prediction of runoff at the dam and compare the effectiveness of different learning rules and activation function. Radial basis function neural network is having a feed-forward structure consisting of hidden layer for a given number of locally tuned units which are fully interconnected to an output layer of linear units.

At a time when global climatic change would seem to be increasing the risk of historically unprecedented changes in

river regimes, it would appear to be appropriate that alternative representations for flood forecasting should be considered.

II. METHODOLOGY

In this study different parameters like learning rule and activation function are employed for rainfall-runoff modeling using Radial basis function neural network model of artificial neural network.

Radial basis functions networks have a very strong mathematical foundation rooted in regularization theory for solving ill-conditioned problems.

The mapping function of a radial basis function network, is built up of Gaussians rather than sigmoids as in MLP networks. Learning in RBF network is carried out in two phases: first for the hidden layer, and then for the output layer. The hidden layer is self-organising; its parameters depend on the distribution of the inputs, not on the mapping from the input to the output. The output layer, on the other hand, uses supervised learning (gradient or linear regression) to set its parameters.

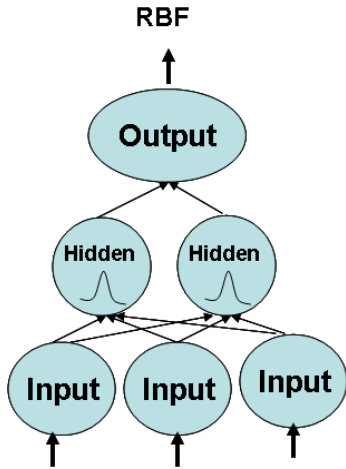


Figure 1. The Radial basis function neural network

In this study we applied different learning rules to the RBF neural network and studied the optimum performance with different activation function. We applied Momentum, Deltabar Delta, Levenberg Marquardt, Conjugate Gradient, Quick prop learning rule with activation function Tanh, Linear Tanh, Sigmoid and Linear Sigmoid.

Performance Measures:

The learning and generalization ability of the estimated NN model is assessed on the basis of important performance measures such as MSE (Mean Square Error), NMSE (Normalized Mean Square Error) and r (Correlation coefficient)

MSE (Mean Square Error):

The formula for the mean square error is:

$$MSE = \frac{\sum_{j=0}^P \sum_{i=0}^N (d_{ij} - y_{ij})^2}{NP} \quad \dots (1)$$

Where

P = number of output PEs,

N = number of exemplars in the data set,

y_{ij} = network output for exemplar i at PE j,

d_{ij} = desired output for exemplar i at PE j.

NMSE (Normalized Mean Square Error):

The normalized mean squared error is defined by the following formula:

$$NMSE = \frac{P N MSE}{\sum_{j=0}^P \frac{N \sum_{i=0}^N d_{ij}^2 - \left(\sum_{i=0}^N d_{ij} \right)^2}{N}} \quad \dots (2)$$

Where

P = number of output processing elements,

N = number of exemplars in the data set,

MSE = mean square error,

d_{ij} = desired output for exemplar i at processing element j.

r (correlation coefficient):

The size of the mean square error (MSE) can be used to determine how well the network output fits the desired output, but it doesn't necessarily reflect whether the two sets of data move in the same direction. For instance, by simply scaling the network output, the MSE can be changed without changing the directionality of the data. The correlation coefficient (r) solves this problem. By definition, the correlation coefficient between a network output x and a desired output d is:

$$r = \frac{\sum_i (x_i - \bar{x})(d_i - \bar{d})}{\sqrt{\sum_i (d_i - \bar{d})^2} \sqrt{\sum_i (x_i - \bar{x})^2}} \quad \dots (3)$$

$$\text{where } \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad \bar{d} = \frac{1}{N} \sum_{i=1}^N d_i$$

The correlation coefficient is confined to the range [-1, 1]. When $r = 1$ there is a perfect positive linear correlation between x and d , that is, they co-vary, which means that they vary by the same amount.

III. STUDY AREA AND DATA SET

The Upper Wardha catchment area lies directly in the path of depression movements which originates in the Bay of Bengal. When the low pressure area is formed in the Bay of Bengal and cyclone moves in North West directions, many times this catchment receives very heavy intense cyclonic precipitation for a day or two. Occurrence of such events have been observed in the months of August and September. Rainfall is so intense that immediately flash runoff, causing heavy flood has been very common feature in this catchment.

For such flashy type of catchment and wide variety in topography, runoff at dam is still complicated to predict. The conventional methods also display chaotic result. Thus ANN based model is built to predict the total runoff from rainfall in Upper Wardha catchment area for controlling water level of the dam.

In the initial reaches, near its origin catchment area is hilly and covered with forest. The latter portion of the river lies almost in plain with wide valleys.

The catchment area up to dam site is 4302 sq. km. At dam site the river has wide fan shaped catchment area which has large variation with respect to slope, soil and vegetation cover.



Figure 2- Location of Upper Wardha dam on Indian map

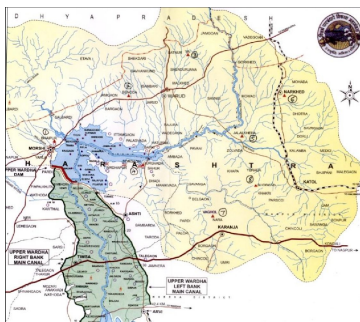


Figure 3- The Wardha river catchment

Data: Rainfall runoff data for this study is taken from the Wardha river catchment area which contains a mix of urban and rural land. The catchments is evenly distributed in eight zones based on the amount of rainfall and geographical survey. The model is developed using historical rainfall runoff data, provided by Upper Wardha Dam Division Amravati, department of irrigation Govt. of Maharashtra. Network is trained by rainfall information gathered from eight telemetric rain-gauge stations distributed evenly throughout the catchment area and runoff at the dam site.

The data is received at the central control room online through this system on hourly basis. The Upper Wardha dam reservoir operations are also fully automated. The amount of inflow, amount of discharge is also recorded on hourly basis. From the inflow and discharge data the cumulative inflow is calculated. The following features are identified for the modeling the neural network.

TABLE I - THE PARAMETERS USED FOR TRAINING THE NETWORK

Month	RG1	RG2	RG3	RG4	RG5	RG6	RG7	RG8	CIF
-------	-----	-----	-----	-----	-----	-----	-----	-----	-----

- Month – The month of rainfall
- Rain1 to Rain8 – Eight rain gauging stations.
- Cum Inflow – Cumulative inflow in dam

Seven years of data on hourly basis from 2001 to 2007 is used. It has been found that major rain fall (90%) occurs in the month of June to October Mostly all other months are dry hence data from five months. June to October is used to train the network

IV. RESULT

The different structures of neural network are employed to learn the unknown characterization of the system from the dataset presented to it. The dataset is partitioned into three categories, namely training, cross validation and test. The idea behind this is that the estimated NN model should be tested against the dataset that was never presented to it before. This is necessary to ensure the generalization. An experiment is performed at least twenty five times with different random initializations of the connection weights in order to improve generalization.

The data set is divided in to training, testing and cross validation data and the network is trained for all models of Radial basis function neural network for 5000 epochs.

The performance results obtain on parameters by applying learning rules Momentum, Deltabar Delta, Levenberg Marquardt, Conjugate Gradient, Quick prop with activation function Tanh, Linear Tanh, Sigmoid, Linear Sigmoid are listed in Table II through Table VI.

TABLE II - RBF NETWORK PERFORMANCE WITH MOMENTUM LEARNING RULE

Parameter 1	MSE	N MSE	Min Abs error	Max Abs error	r
Tanh	0.106	0.124	0.034	0.465	0.534
Linear Tanh	0.097	0.105	0.024	0.212	0.639
Sigmoid	0.089	0.093	0.047	0.421	0.678
Linear Sigmoid	0.094	0.132	0.041	0.381	0.689

TABLE III - RBF NETWORK PERFORMANCE WITH DELTAR DELTA LEARNING RULE

Parameter 2	MSE	N MSE	Min Abs error	Max Abs error	r
Tanh	0.093	0.141	0.051	0.564	0.651
Linear Tanh	0.190	0.241	0.041	0.412	0.591
Sigmoid	0.143	0.215	0.032	0.495	0.543
Linear Sigmoid	0.086	0.095	0.067	0.315	0.603

TABLE IV - RBF NETWORK PERFORMANCE WITH L. M. LEARNING RULE

Parameter 3	MSE	N MSE	Min Abs error	Max Abs error	r
Tanh	0.076	0.064	0.018	0.143	0.854
Linear Tanh	0.086	0.094	0.028	0.298	0.732
Sigmoid	0.083	0.094	0.020	0.228	0.634
Linear Sigmoid	0.089	0.095	0.034	0.469	0.758

TABLE V - RBF NETWORK PERFORMANCE WITH CONJUGATE GRADIENT LEARNING RULE

Parameter 4	MSE	N MSE	Min Abs error	Max Abs error	r
Tanh	0.094	0.165	0.051	0.312	0.646
Linear Tanh	0.089	0.094	0.059	0.215	0.633
Sigmoid	0.092	0.134	0.041	0.474	0.701
Linear Sigmoid	0.094	0.124	0.064	0.541	0.732

TABLE VI - RBF NETWORK PERFORMANCE WITH QUICK PROP. LEARNING RULE

Parameter 5	MSE	N MSE	Min Abs error	Max Abs error	r
Tanh	0.133	0.245	0.042	0.465	0.584
Linear Tanh	0.169	0.212	0.054	0.514	0.601
Sigmoid	0.106	0.256	0.059	0.329	0.563
Linear Sigmoid	0.098	0.112	0.046	0.311	0.609

The parameters and performance for RBF model with different learning rule and activation function are compared on the performance scale and are listed in the Table VII shown below. The comparative analysis of the MSE and r (the correlation coefficient) is done.

TABLE VII – COMPARISON OF PERFORMANCE PARAMETERS

		Tanh		Linear Tanh		Sigmoid		Linear Sigmoid	
		MSE	r	MSE	r	MSE	r	MSE	r
1	Momentum	0.106	0.534	0.0975	0.639	0.089	0.678	0.094	0.689
2	Deltabar Delta	0.0931	0.651	0.1906	0.591	0.143	0.543	0.086	0.603
3	L.M.	0.07629	0.854	0.0861	0.732	0.083	0.634	0.0894	0.758
4	Conjugate Gradient	0.0946	0.646	0.0894	0.633	0.0921	0.701	0.0945	0.732
5	Quickprop	0.1331	0.584	0.1691	0.601	0.106	0.563	0.0986	0.609

After training the network the optimum performance is studied and it is found that Levenberg Marquardt learning rule and Tanh activation function produce optimal result. In the Table-VIII the parameters and the best performances for Radial basis function neural network are listed.

TABLE VIII- RBF NETWORK PARAMETERS

Parameter	Performance
MSE	0.07629
NMSE	0.06431
Min Abs Error	0.01943
Max Abs Error	0.14387
r	0.85437

Fig 4 shows the plot of actual Vs predicted optimum values for Radial basis function neural network found with Levenberg Marquardt learning rule and Tanh activation function.

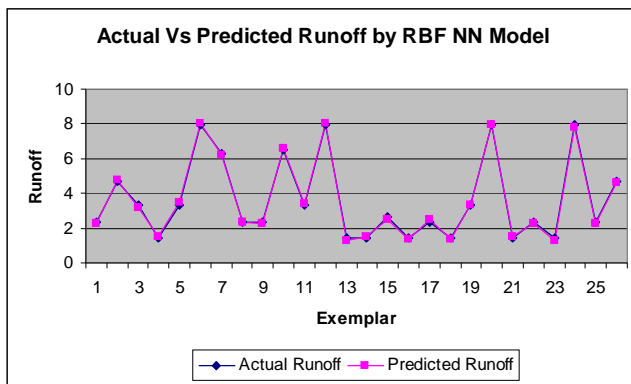


Figure 4.– Actual Vs. Predicted runoff by RBF for L.M. and Tanh

The error found in the actual and predicted runoff at the dam site is plotted for RBF network as shown in the Figure 5.

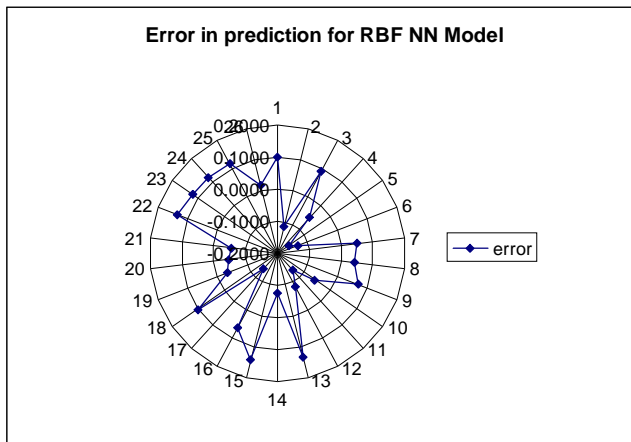


Fig 5 – Error graph of RBF Model for L.M. and Tanh

The main advantage of RBF is that it finds the input to output map using local approximators. Each one of these local pieces is weighted linearly at the output of the network. Since they

have fewer weights, these networks train extremely fast and require fewer training samples.

V. CONCLUSION

An ANN-based short-term runoff forecasting system is developed in this work. A comparison between five different learning rules with four activation function for optimal performance for Radial basis function neural network model is made. We find that Radial basis function neural network with Levenberg Marquardt learning rule and Tanh activation function is more versatile than other approaches studied. Radial basis function neural network with Levenberg Marquardt learning rule and Tanh activation function is performing better as compare to other approaches studied as far as the overall performance is concerned for forecasting runoff for 3 hrs lead time. Other approaches studied are also performing optimally. Which means that static model of Radial basis function neural network with Levenberg Marquardt learning rule and Tanh activation function is powerful tool for short term runoff forecasting for Wardha River basin.

ACKNOWLEDGMENT

This study is supported by Upper Wardha Dam Division Amravati, department of irrigation Govt. of Maharashtra, India

REFERENCES

- [1] P. Srivastava, J. N. McVair, and T. E. Johnson, "Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed," *Journal of the American Water Resources Association*, vol. 42, pp. 545-563, Jun 2006.
- [2] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, pp. 359-366, 1989.
- [3] M. C. Demirel, A. Venancio, and E. Kahya, "Flow forecast by SWAT model and ANN in Pracana basin, Portugal," *Advances in Engineering Software*, vol. 40, pp. 467-473, Jul 2009.
- [4] A. S. Tokar and M. Markus, "Precipitation-Runoff Modeling Using Artificial Neural Networks and Conceptual Models," *Journal of Hydrologic Engineering*, vol. 5, pp. 156-161, 2000.
- [5] S. Q. Zhou, X. Liang, J. Chen, and P. Gong, "An assessment of the VIC-3L hydrological model for the Yangtze River basin based on remote sensing: a case study of the Baohe River basin," *Canadian Journal of Remote Sensing*, vol. 30, pp. 840-853, Oct 2004.
- [6] R. J. Zhao, "The Xinanjiang Model," in *Hydrological Forecasting Proceedings Oxford Symposium*, IASH, Oxford, 1980 pp. 351-356.
- [7] R. J. Zhao, "The Xinanjiang Model Applied in China," *Journal of Hydrology*, vol. 135, pp. 371-381, Jul 1992.
- [8] D. Zhang and Z. Wanchang, "Distributed hydrological modeling study with the dynamic water yielding mechanism and RS/GIS techniques," in *Proc. of SPIE*, 2006, pp. 63591MI-12.
- [9] J. E. Nash and I. V. Sutcliffe, "River flow forecasting through conceptual models," *Journal of Hydrology*, vol. 273, pp. 282-290, 1970.
- [10] D. Zhang, "Study of Distributed Hydrological Model with the Dynamic Integration of Infiltration Excess and Saturated Excess Water Yielding Mechanism." vol. Doctor Nanjing: Nanjing University, 2006, p. 190-529

- [11] E. Kahya and J. A. Dracup, "U.S. Streamflow Patterns in Relation to the El Niño/Southern Oscillation," *Water Resour. Res.*, vol. 29, pp. 2491-2503, 1993.
- [12] K. J. Beven and M. J. Kirkby, "A physically based variable contributing area model of basin hydrology," *Hydrological Science Bulletin*, vol. 43, pp. 43-69, 1979.
- [13] N. J. de Vos, T. H. M. Rientjes, "Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation", *Hydrology and Earth System Sciences*, European Geosciences Union, 2005, 9, pp. 111-126.
- [14] Holger R. Maier, Graeme C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications", *Environmental Modelling & Software*, ELSEVIER, 2000, 15, pp. 101-124.
- [15] T. Hu, P. Yuan, etc. "Applications of artificial neural network to hydrology and water resources", *Advances in Water Science*, NHRI, 1995, 1, pp. 76-82.
- [16] Q. Ju, Z. Hao, etc. "Hydrologic simulations with artificial neural networks", *Proceedings-Third International Conference on Natural Computation*, ICNC, 2007, pp. 22-27.
- [17] G. WANG, M. ZHOU, etc. "Improved version of BTOPMC model and its application in event-based hydrologic simulations", *Journal of Geographical Sciences*, Springer, 2007, 2, pp. 73-84.
- [18] K. Beven, M. Kirkby, "A physically based, variable contributing area model of basin hydrology", *Hydrological Sciences Bulletin*, Springer, 1979, 1, pp. 43-69.
- [19] K. Thirumalaiah, and C.D. Makarand, *Hydrological Forecasting Using Neural Networks Journal of Hydrologic Engineering*. Vol. 5, pp. 180-189, 2000.
- [20] G. WANG, M. ZHOU, etc. "Improved version of BTOPMC model and its application in event-based hydrologic simulations", *Journal of Geographical Sciences*, Springer, 2007, 2, pp. 73-84.
- [21] H. Goto, Y. Hasegawa, and M. Tanaka, "Efficient Scheduling Focusing on the Duality of MPL Representatives," *Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07)*, IEEE Press, Dec. 2007, pp. 57-64.
- [22] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, "Artificial neural networks in hydrology I: preliminary concepts", *Journal of Hydrologic Engineering*, 5(2), pp. 115-123, 2000



Rahul Deshmukh received the B.E. and M.E. degrees in Electronics Engineering from Amravati University. During 1996-2007, he stayed in Government College of Engineering, Amravati in department of Electronics and telecommunication teaching undergraduate and postgraduate students. From 2007 till now he is with Indian Institute of Technology (IIT) Bombay, Mumbai. His area of research are artificial intelligence and neural networks.



A. A. Ghatol received the B.E. from Nagpur university followed by M. Tech and P.h.d. from IIT Bombay. He is best teacher award recipient of government of Maharashtra state. He has worked as director of College of Engineering Poona and Vice-Chancellor, Dr. Babasaheb Ambedkar Technological University, Lonere, Raigad, India. His area of research is artificial intelligence, neural networks and semiconductors.

Analysis of impact of Symmetric Encryption Algorithms in Data Security Model of Grid Networks

N. Thenmozhi

Department of Computer Science
N.K.R. Govt. Arts College for Women
Namakkal-637 001, India.

M. Madheswaran

Department of Electronics and Communication Engg.,
Muthayammal Engineering College
Rasipuram-637 408, India.

Abstract—The symmetric and asymmetric encryption algorithms are commonly used in grid software to provide necessary security. The use of symmetric encryption algorithm will significantly affect the network communication performance.

In this paper, the impact of using different popular and commonly used symmetric key cryptography algorithms for encrypting data in a typical grid commuting environment is analyzed. It is obvious that the use of encryption and decryption at application layer will certainly have an impact in the application layer performance in terms of speed. In this work, we have studied its impact at network layer performance in a typical grid computing environment in the algorithms such as DES, Triple DES, AES, Blow Fish, RC2 and RC6. The performances are measured through simulation studies on ns2 by simulating these algorithms in GARUDA Grid Network Topology.

Keywords— Grid Security; Encryption; ECGIN; ERNET; GARUDA; PPlive; GridFTP;

I. INTRODUCTION

Internet and Grid computing applications are growing very fast, so the needs to protect such applications have increased. Encryption algorithms play a main role in information security systems. On the other side, those algorithms consume a significant amount of computing resources such as CPU time, memory, and battery power.

The Globus Toolkit is the very commonly used software for Grid computing. It provides different kinds of security for grid computing. The Grid Security Infrastructure (GSI) of Globus and a Public Key Infrastructure (PKI) provide the technical framework (including protocols, services, and standards) to support grid computing with five security capabilities: user authentication, data confidentiality, data integrity, non-repudiation, and key management.

A. Security Issues

Authentication is the process of verifying the validity of a claimed individual and identifying who he or she is. Authentication is not limited to human beings; services,

applications, and other entities may be required to authenticate also. Basic authentication is the simplest web-based authentication scheme that works by sending the username and password within the request. Generally authentication is achieved through the presentation of some token that cannot be stolen (forged). This can be either peer-to-peer relationship (password for client and server) or through a trusted third party (certification authority or Kerberos server). Biometrics characteristics can also be used to a service for authentication purpose, since a unique identification of human being can give more security for example a finger print scanner can be used to log into a local machines. Trust can be defined as the assured reliance on the character, ability, strength, or truth of someone or something.

Access Control is the ability to limit and control the access to host systems and applications via communications links. The process of authorization is often used as a synonym for access control, but it also includes granting the access or rights to perform some actions based on access rights.

Data integrity assures that the data is not altered or destroyed in an unauthorized manner. Integrity checks are provided primarily via hash functions (or “message digests”). Data confidentiality, Sensitive information must not be revealed to parties that it was not meant for. Data confidentiality is often also referred to as privacy. The standard approach to ensure confidentiality is through encryption, which is the application of an algorithm that transforms “plaintext” to “cipher text” whose meaning is hidden but can be restored to the original plaintext by another Algorithm (the invocation of which is called decryption).

Key management deals with the secure generation, distribution, authentication, and storage of keys used in cryptography. Nonrepudiation refers to the inability of something that performed a particular action such as a financial transaction to later deny that they were indeed responsible for the event.

Basically, security requires at least three fundamental services: authentication, authorization, and encryption. A grid resource must be authenticated before any checks can be done as to whether or not any requested access or operation is

allowed within the grid. Once the grid resources have been authenticated within the grid, the grid user can be granted certain rights to access a grid resource. This, however, does not prevent data in transit between grid resources from being captured, spoofed, or altered [18]. The security service to insure that this does not happen is encryption. Obviously, use of data encryption certainly will have its impact on application layer performance. But, in this work we will examine its impact on total network performance. In this paper, we will study the impact of four symmetric encryption algorithms in a typical grid network.

The use of cryptography will certainly have an impact on network performance in one way or another. So we decided to model an application layer encryption-decryption scenario in a typical grid computing environment and study its impact on network performance through network simulations.

B. Security Methods

Symmetric encryption: Using the same secret key to provide encryption and decryption of data. Symmetric cryptography is also known as secret-key cryptography.

Asymmetric encryption: Using two different keys for encryption and decryption. The public key encryption technique is the primary example of this using a "public key" and a "private key" pair. So it is referred as public-key cryptography.

Secure Socket Layer/Transport Layer Security (SSL/TLS): These are essentially the same protocol, but are referred to one another differently. TLS has been renamed by the IETF, but they are based on the same RFC.

Public Key Infrastructure (PKI): The different components, technologies, and protocols that make up a PKI environment. Grid security implementations are predominantly built on public key infrastructure (PKI) (Housely et al., 2002; Tuecke et al., 2004). In a PKI each entity (e.g. user, service) possesses a set of credentials comprised of a cryptographic key and a certificate.

Mutual Authentication: Instead of using an Lightweight Distribution Access Protocol (LDAP) repository to hold the public key (PKI), two parties who want to communicate with one another use their public key stored in their digital certificate to authenticate with one another.

C. The symmetric key Encryption Algorithms

Data Encryption Standard(DES), was the first encryption standard to be recommended by NIST (National Institute of Standards and Technology). It is based on the IBM proposed algorithm called Lucifer. DES became a standard in 1974. Since that time, many attacks and methods were recorded that exploit the weaknesses of DES, which made it an insecure block cipher[22].

Advanced Encryption Standard(AES), is the new encryption standard recommended by NIST to replace DES. Rijndael (pronounced Rain Doll) algorithm was selected in

1997 after a competition to select the best encryption standard. Brute force attack is the only effective attack known against it, in which the attacker tries to test all the characters combinations to unlock the encryption. Both AES and DES are block ciphers[20].

Blowfish is a variable length key, the block size is 64 bits, and the key can be any length up to 448 bits block cipher. This algorithm can be optimized in hardware applications though it's mostly used in software applications. Though it suffers from weak keys problem, no attack is known to be successful against [8][23].

RC2 is a block cipher with a 64-bits block cipher with a variable key size that range from 8 to 128 bits. RC2 is vulnerable to a related-key attack using 234 chosen plaintexts [20].

Authentication and authorization has been a basic and necessary Service for internet transactions. Several new standards have merged which allow dynamic access control based on exchanging user attributes. Unfortunately, while providing highly secure and flexible access mechanisms are a very demanding task. Authentication and Authorization Infrastructures (AAIs) can provide such integrated federations of security services. They could, in particular, provide attribute based access control (ABAC) mechanisms and mediate customers' demand for privacy and vendors' needs for information [10].

II. LITERATURE SURVEY

The Globus Security Infrastructure (GSI) is one of the most famous security architecture. GSI is based on Public Key Infrastructure (PKI), which performs mutual authentication via X.509 certificates. The author describes present a password-based grid security infrastructure (PBGSI), which authenticates clients by authenticated key exchange (AuthA) methods and uses improved Chaffing and Winnowing for secure data transfer. By using password-based methods in authentication, authorization and delegation, PBGSI provides convenient interface for the user. At the same time, encryption-less secure data transfer improves the performance; and mechanisms used in our scheme (time-stamp etc.) enhance the security of the whole grid [11].

A grid environment is built to verify the feasibility and the efficiency of the extended OCSP protocol. The paper deals with the running requirement and the data description of the client and each extended OCSP responder in detail. It describes the processing algorithm of each responder. In order to improve the efficiency of the system, the path length constraint and time constraint of request transmitting are designed specially. Theory and experiments all prove that the extended OCSP system improves the efficiency of certificate verification effectively [12].

Recently, Authentication protocol has been recognized as an important factor for grid computing security. This paper [20] described a new simple and efficient Grid authentication system providing user anonymity. It is based on hash function, and mobile users only do symmetric encryption and

decryption and it takes only one round of messages exchange between the mobile user and the visited network, and one round of message exchange between the visited network and the corresponding home network.

There are number of projects investigating attribute-based authentication such as the VO Privilege Project, GridShib, and PERMIS. However, there are quite a few decision dimensions when it comes about designing this scheme in grid computing [10].

Authentication in the grid environment can be performed in two ways either in the application layer part or in the communication part. Cryptography plays a major role to implement authentication. It is obvious that the use of encryption and decryption at application layer will certainly have an impact in the application layer performance in the grid environment. In this paper, we have simulated the encryption algorithms in a typical grid network scenario using the results from the paper [1].

A. Europe-China Grid Internetworking (EC-GIN) Project

The Internet communication infrastructure (the TCP/IP protocol stack) is designed for broad use; as such, it does not take the specific characteristics of Grid applications into account. This one-size-fits-all approach works for a number of application domains, however, it is far from being optimal general network mechanisms, while useful for the Grid, and cannot be as efficient as customized solutions. While the Grid is slowly emerging, its network infrastructure is still in its infancy. Thus, based on a number of properties that make Grids unique from the network perspective, the project EC-GIN (Europe-China Grid Internetworking) will develop tailored network technology in dedicated support of Grid applications. These technical solutions will be supplemented with a secure and incentive-based Grid Services network traffic management system, which will balance the conflicting performance demand and the economic use of resources in the network and within the Grid [30].

By collaboration between European and Chinese partners, EC-GIN parallels previous efforts for real-time multimedia transmission across the Internet: much like the Grid, these applications have special network requirements and show a special behavior from the network perspective.

B. The ERNET Project

ERNET[26] (Education and Research Network) was the first dedicated and integrated step taken towards to enable the research and education community in India to leverage the benefits of ICTs. ERNET India aims at developing, setting up and operating nationwide state-of-the-art computer communication infrastructure and providing services to the users in academic and research institutions, Government organizations, and industry, in line with technology developments and national priorities. Dissemination, training and knowledge transfer in the field of computer communication and information technology are an integrating part of ERNET mission.

ERNET also acts as a bridge for co-operation with other countries in the area of computer communications, information technology, computer networking and other related emerging technologies.

The ERNET network has 15 Points of Presence spread throughout India serving 1389 institutions, including 152 universities, 284 agricultural universities and many other research organizations. It has 14 points of peering for Internet bandwidth connectivity using submarine cables.

The network comprises a mix of terrestrial and satellite-based wide area networks. It provides a wide range of operation and application services. As of today, universities, academic institutions, R&D labs and schools, etc. use ERNET for a variety of applications and services including email, file transfer, database access, world wide web, web hosting, mail relaying, security solutions, distant learning and grids.

ERNET is the first network in the country to provide dual stack access of Internet protocol version 6 (IPv6) and Internet protocol version 4 (IPv4) test beds to its users to develop, test and implement IPv6 based mail, Domain name Services, Web applications and products.

ERNET has deployed many overlay networks over its terrestrial and satellite network under different schemes. Some examples are GARUDA (see below), UGC-Infonet, interconnecting Indian universities, ICAR-Net, interconnecting Agricultural Research centers, Universities and Stations, and several pilot projects aiming at interconnecting schools. Separate networks were implemented to allow DAE institutes to connect to the GÉANT network and to participate in LHC activities.

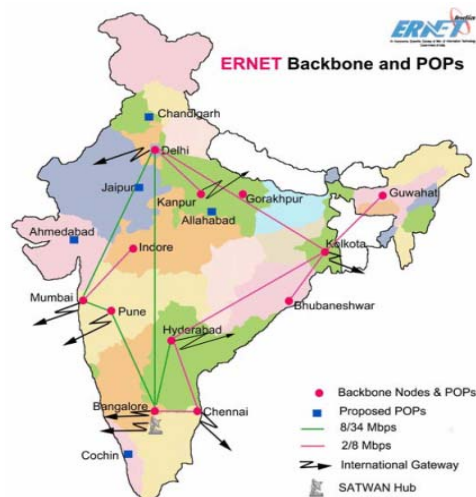


Figure 1. The ERNET Topology [18]

C. Overview of GARUDA Project

GARUDA[27] initiative is a collaboration of science researchers and experimenters on a nation-wide grid of computational nodes, mass storage and scientific instruments that aims to provide the technological advances required to

enable data and compute intensive science of the 21st century. One of GARUDA's most important challenges is to strike the right balance between research and the daunting task of deploying that innovation into some of the most complex scientific and engineering endeavours being undertaken today.

The Department of Information Technology (DIT) has funded the Center for Development of Advanced Computing (C-DAC[27]) to deploy the nation-wide computational grid 'GARUDA' which today connects 45 institutions across 17 cities in its Proof of Concept (PoC) phase with an aim to bring "Grid" networked computing to research labs and industry. In pursuit of scientific and technological excellence, GARUDA PoC has also brought together the critical mass of well-established researchers.

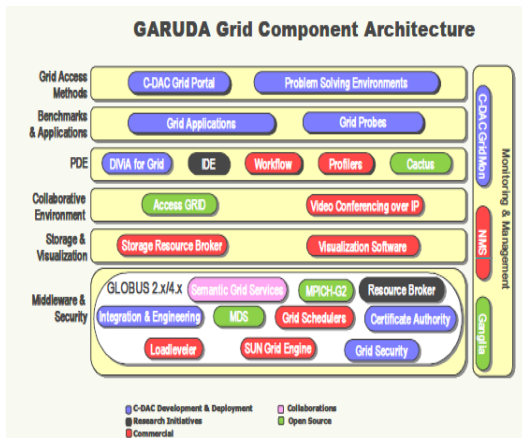


Figure 2. GARUDA Grid Component Architecture [29]

C. Present Network Architecture

The GARUDA network is a Layer 2/3 MPLS Virtual Private Network [VPN] connecting selected institutions at 10/100 Mbps with stringent quality and Service Level Agreements. The network has been contracted as a man-aged service with a peak capacity of 2.43 Gbps across 17 cities. This network is a pre-cursor to the next generation Gigabit speed nation-wide Wide Area Network with high performance computing resources and scientific instruments for seamless collaborative research and experiments. The PoC network was established at all the GARUDA partner institutes in close collaboration with ERNET who are responsible for the operation, maintenance and management of this network.

D. Computational Resources in GARUDA

In this collaborative grid project, various resources such as high performance computing systems (HPC) and satellite based communication systems have been committed by different centers of C-DAC and GARUDA partners. It may be noted that since the resources are diverse in nature, one of the major challenges of GARUDA is to deploy appropriate tools and middleware to enable applications to run seamlessly across the grid.

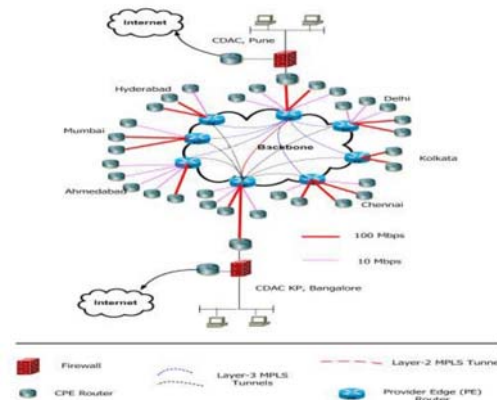


Figure 3. GARUDA topology - EU-INDIA GRID [18]

E. Network Simulator

The Grid Computing paradigm has been widely adopted within the research community for scientific computing. Grid Computing is used as a method by which access is seamlessly given to a set of heterogeneous computational resources across a dynamic set of physical organizations, supplying massive computing and storage capabilities. Within a Grid environment, computational jobs are submitted to and run on suitable resources and data is stored and transferred transparently without knowing its geographic location. All of this behavior will obviously show its impact on the underlying network infrastructure and the data generated within a Grid environment may substantially affect the network performance due to the volume involved.

We will use NS2 to simulate the network, but it is well known that NS2 doesn't implement any security features. Till now, there is no option for simulating security things in NS2. The reasons for lack of security features in ns2 are:

- Security is a subtle thing related to many aspects, which is much different from other kinds of network protocols.
- Generally there will not be any real data or packet to encrypt or decrypt in ns2.
- The scope of a simulation will be minimizing the overall simulation time. But if we do real encryption or decryption in simulator, then it will go beyond the concept of a simulator.
- Lack of support for sending real payload in ns2.
- Lack of support for handling socket connection like real TCP/IP scenario.
- Ns2 simulator has limitation in simulating simultaneous threaded processes to mimic real socket connections.

Ns2 [16] is an object oriented simulator, written in C++, with an OTCL interpreter as a frontend. The simulator supports a class hierarchy in C++, and a similar class hierarchy within the OTcl interpreter. The root of this

hierarchy is the class TcObject. Users create new simulator objects through the interpreter. Applications sit on top of transport agents in ns and there are two basic types of applications: traffic generators and simulated applications. Currently, there are four C++ classes derived from the traffic generator class [20]. Traffic Generator: EXPOO_Traffic, POO_Traffic, CBR_Traffic, TrafficTrace.

However, none of these classes match the traffic characteristics of PPLive, and of GridFTP. So we decided to simulate encryption in ns2 at application layer, by modeling a new encrypted traffic generator.

III. MODELING GRID AND GRID TRAFFIC IN NS2

Though there are different kinds of security requirements or models for grid computing systems, the role of a symmetric key encryption algorithm and its impact will be a significant one when implemented in application layer that will affect the performance in terms of time. In this work, we have simulated the workload of different Symmetric Key Encryption algorithms such as DES, Triple DES, AES, Blow Fish, RC2 and RC6 at application layer using Network Simulator tool. The proposed traffic model is based on the model used in ECGIN for symmetric key encryption and GridFTP as a cross traffic. The proposed model is implemented on the Indian grid network topology GARUDA, to study the impact of the encryption based traffic model.

A. Modeling Encrypted PPLive Traffic

Along with the rapid development of P2P file sharing and IPTV video services, P2P streaming services have become a core multi-user video sharing application on the Internet. The focus of grid technology in the video area is generally on the resource scheduling and replica management aspects, while the service traffic characteristics are still similar to the traditional video service. In depth work has already been carried out in the areas of monitoring and modeling video traffic[25]. Therefore, exploring the developing trends of grid systems, video sharing, monitoring and the analysis of P2P IPTV traffic are interesting and promising topics of research.

The time interval between two packets and the size of each packet waiting for sending out is very important when modeling actual traffic. Therefore if the model can accurately match these two characteristics, it can be said to generate traffic that is similar to the actual data. The EC-GIN project built a new traffic generator to model the actual traffic called Lognormal Traffic, which is primarily responsible for controlling the packets time interval and the packet sizes.

In this work, we extended the traffic model of PPLive (Lognormal Traffic) to support a simulated encryption-decryption scenario.

Based on traffic model of EC-GIN, an algorithm has been put forward to control the packet generation sequence. First, data initialization is performed as follows:

- Send a video packet when simulation begins.

- Compute the next video packet sending time. Put it into a variable NextT.

Next, the time needed to send the next packet is computed. To account for different packet sizes, different parameters are used to calculate inter-video packet time (variable NextT) and the inter-control packet time (array t_i). The values of t_1 to t_n are summed to variable SmallT. As long as the value of SmallT is less than NextT, t_i is used as the inter- packet time for sending small packets (control packets). Otherwise, a large packet(video packet) is sent immediately with an inter-packet time of NextT - (SmallT - t_i).

In addition to the above process, we have delayed the packet transmission with respect to the size of the packet to be sent and the selected encryption algorithm.

So the new Scheduled Transmission Time will be equal to the sum of inter-packet time and the time taken for encrypting the packet by the selected algorithm.

In our implementation we have simulated the encryption algorithms in a typical grid network scenario just by including the encryption delay at the traffic generator using the results from the paper [1]. In the traffic model of ECGIN, they used UDP in their design. We have decided to use TCP in our design, because, TCP is the most commonly used transport protocol in grid network communication.

B. Modeling GridFTP

The GridFTP tool of Globus Toolkit is one of the most important components provided by Globus for moving large amounts of data in bulk. GridFTP is based on FTP, the highly- popular Internet file transfer protocol. Given the characteristics of Grid traffic - often a mixture of short, sporadic service calls and bulk data transfers - a GridFTP simulation scenario differs from other traffic models and is therefore important for testing Grid-specific network mechanisms. The GridFTP simulator of EC-GEN was developed with the OTCL language to mimic this GridFTP traffic. The EC-GEN GridFTP is embedded in a gridftp.tcl file. In this work we just used GridFTP as a background cross traffic during evaluation the impact of encrypted PPLive traffic. The three major parameters defined for the GridFTP simulator are:

- Bandwidth: this parameter is used to set the total bandwidth of the link. By default, this parameter is set to 1.0Mbps. With this and the ratio parameter, we can determine the "rate_" parameter for each FTP instance.
- Parallel: this parameter is used to set the parallel GridFTP streams. By default, this is set to 4. Since each GridFTP stream can be simulated by FTP, this parameter will actually set the number of FTP instances for the GridFTP simulator.
- Ratio: this parameter is used to set the throughput ratio among the parallel streams. By default, this is set to 1:1:1:1 which means each stream will transmit packets at an equal speed.

The GridFTP simulator consists of two classes. One is the GridFTP class and the other is the GridFTPSink class. We also override two methods for the basic Simulator class, attach-agent and connect, with which the GridFTP instance can be attached to the network node and be connected to the GridFTPSink instance.

C. The Simulation of GARUDA Network in ns2

The following NAM (Network Animator) output shows the model of GARUDA network simulated on ns2. The topology was derived from the information provided by the ERNET and GARUDA projects [26][27].

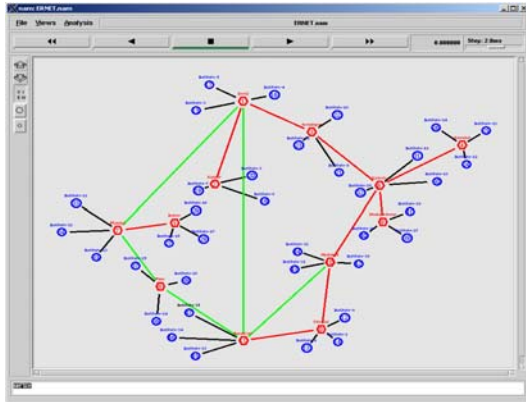


Figure 5. The Simulated GARUDA Topology

- The links shown in green are 8/34Mbps links
- The links shown in red are 2/8 Mbps links
- Nodes shown as red hexagon are backbones and POPs
- Nodes shown as blue circles are the connected institutes

IV. SIMULATION RESULTS AND DISCUSSION

A simple model of GARUDA grid network has been simulated in ns2 and the impact of different encryption schemes on network performance has been evaluated. A normal 2 GHz Pentium IV computer with 1 GB RAM was used for this simulation.

A. Traffic models

In order to create the different traffic scenarios files we used different types of grid traffics mentioned in ECGIN project. They are GridFTP Traffic and PPLive Traffic.

Some of the simulation parameters are

Number of Backbone and POP nodes	12
Number of Simulated Institution Nodes	36
Routing Protocol	DV
Backbone Link Capacity	8/34 Mbps

Institution to Backbone Links

Queue Type

2/8 Mbps

DropTail

We have simulated a encrypted PPLive traffic from one node to another (in this topology, from Madras to Delhi) and used some GridFTP cross traffic.

B. Performance

The following graph shows the performance of the network with respect to different cryptography algorithms used in application layer.

The Throughput

The following graphs show the comparison of throughput in different encryption schemes over time.

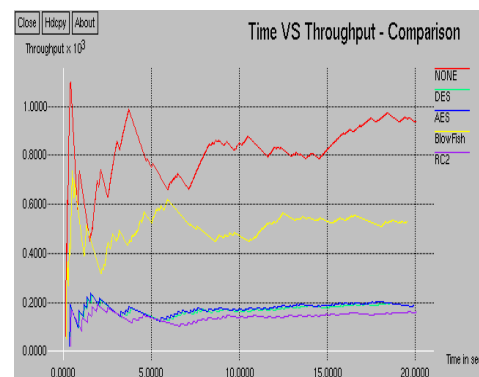


Figure 6. Time VS Throughput – Comparison

The following graph shows the average throughput. The throughput in the case of Blowfish based scheme was good.

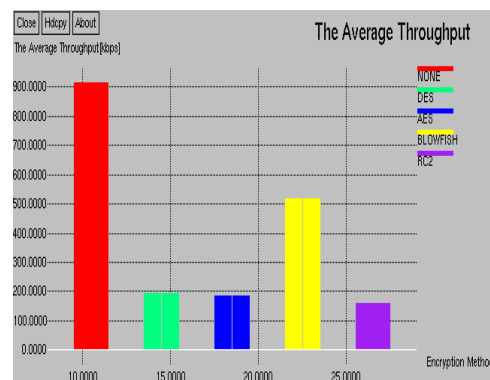


Figure 7. The Average Throughput

The Received Packets comparison

The following graphs show the comparison of time and received packets in different encryption schemes.

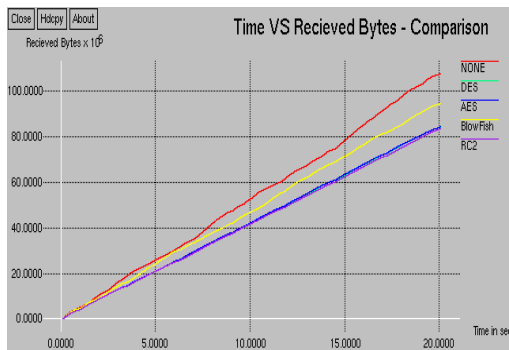


Figure 8. The Time VS Received Packets – Comparison

The End to End Delay

The following graphs show the comparison of end to end delay in different encryption schemes over time

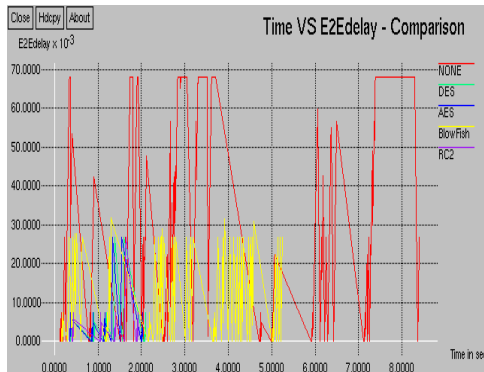


Figure 9. Time VS E2Edelay - Comparison

The Average Delay

The following graphs show the average delay in different encryption schemes.

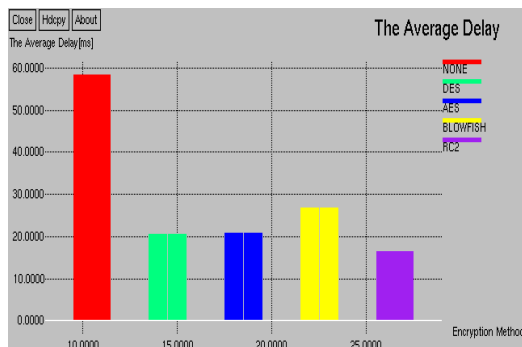


Figure 10. The Average Delay

Even though all the transmitted packets were received successfully, the throughput and delay was much affected by the retransmission of the packets during the packet loss or drop. This retransmission of packet had an impact on throughput. Faster the encryption algorithm, higher the bandwidth it will try to use. So it will increase delay, packet loss as well as drop at intermediate nodes.

V. CONCLUSION

The security is a very important issue in grid network design. Apart from authentication and authorization, the use of symmetric encryption algorithm for grid data security is also having significant impact on the design and performance of grid networks. A model for grid security infrastructure has been implemented on network simulator ns2 and the impact of use of encryption algorithms in network performance has been measured. We have simulated a simplified model of GARUDA grid network in ns2 and simulated some of the basic traffic types of grid network (proposed in ECGIN). As shown in the graphs in previous section, the use of cryptography at application layer has obvious impact on network performance. Depending on the cryptographic algorithms, the delay in delivery of packet is proportional with respect to time. Due to queuing delay at the intermediate node, the faster algorithm provides better throughput with a little bit of delay in packet delivery.

Future works may address the issues of impact of asymmetric encryption algorithms used in a grid network for authentication and other purposes. Further, the work may be extended for implementing some other traffic types of grid network.

REFERENCES

- [1] Diaa Salama Abd Elminaam, Hatem Mohamed Abdul Kader, and Mohiy Mohamed Hadhoud, "Evaluating The Performance of Symmetric Encryption Algorithms" International Journal of Network Security, Vol.10, No.3, PP.216-222
- [2] D. S. Abdul. Elminaam, H. M. Abdul Kader and M. M. Hadhoud, "Performance Evaluation of Symmetric Encryption Algorithms," Communications of the IBIMA Volume 8, 2009 ISSN: 1943-7765.
- [3] Aamer Nadeem, "A Performance Comparison of Data Encryption Algorithms", IEEE 2005.
- [4] Earle, "Wireless Security Handbook,". Auerbach Publications 2005
- [5] Priya Dhawan., "Performance Comparison: Security Design Choices", Microsoft
- [6] Edney, "Real 802.11 Security: Wi-Fi Protected Access and 802.11i", Addison Wesley 2003.
- [7] Hardjono, "Security In Wireless LANS And MANS", Artech House Publishers 2005
- [8] Bruce Schneier, "Applied Cryptography", John Wiley & Sons, Inc 1996
- [9] Ronald L. Rivest, M.J.B. Robshaw, R. Sidney, and Y.L. Yin, "The

- RC6TM, Block Cipher”, Version 1.1 - August 20, 1998.
- [10] Christian Schläger, Manuel Sojer, Björn Muschall, and Günther Pernul
 , “Attribute-Based Authentication and Authorisation Infrastructures for E-Commerce Providers”, K. Bauknecht et al. (Eds.): EC-Web 2006, LNCS 4082, pp. 132 – 141, 2006.
 - [11] Zhun Cai, “A Password-based Grid Security Infrastructure” 10.1109 /ICDS.2008.39, Second International Conference on The Digital Society, Institute of Digital Technology AISINO Inc.
 - [12] Shaomin Zhang, Baoyi Wang, *Hebei Province*, “Research on An Extended OSCP Protocol for Grid”, Proceedings of the 7th World Congress on Intelligent Control and Automation, 25 - 27, 2008, China.
 - [13] Ronghui Wu, Renfa Li, Fei Yu ,guangxue,Cheng Xu, “Research on User Authentication for Grid Computing Security”, Proceedings of the Second International Conference on Semantics, Knowledge, and Grid (SKG'06) 0-7695-2673-X/06 \$20.00 © 2006.
 - [14] Anna Cinzia Squicciarini, Elisa Bertino and Sebastien Goasguen, “Access Control Strategies for Virtualized Environments in Grid Computing Systems”, Proceedings of the 11th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS'07) 0-7695-2810-4/07 \$20.00 © 2007.
 - [15] Marty Humphery, Mary R. Thomson, and Keith R.Jackson, “Security for Grids”, Proceeding of the IEEE, Vol 93, No.3, pp.644-650, March 2005.
 - [16] Europe-China Grid InterNetworking, European Sixth Framework STREP FP6-2006-IST-045256, Deliverable D2.1, Ns2 code for Grid network simulation. The EC-GIN Consortium, Europe-China Grid InterNetworking, Survey of Grid Simulators, Network-level Analysis of Grid Applications, The EC-GIN Consortium.
 - [17] International Technical Support Organization, “Introduction to Grid Computing with Globus”, September 2003, IBM Corporation.
 - [18] <http://partners.euindiaagrid.eu/deliverables/D3.1.html>
 - [19] <http://www.faqs.org/rfcs/rfc2828.html>
 - [20] <http://msdn2.microsoft.com/en-us/library/ms978415.aspx>, Developer Network October 2002.
 - [21] http://en.wikipedia.org/wiki/Block_cipher
 - [22] <http://www.tropsoft.com/strongenc/des.htm>
 - [23] <http://www.eskimo.com/~weidai/benchmarks.html>
 - [24] Coder's Lagoon, <http://www.hotpixel.net/software.html>
 - [25] <http://www.ec-gin.eu>
 - [26] <http://www.eis.ernet.in>
 - [27] www.garudaindia.in
 - [28] <http://www.euindiaagrid.eu/>
 - [29] www.cdac.in
 - [30] http://www.euindiaagrid.eu/index.php/documents/doc_download/11-einfrastructures-across-europe-and-india

Mrs. N. Thenmozhi is working as Assistant Professor, Department of Computer Science in N.K.R. Govt. Arts College for Women, Namakkal. She obtained her Bachelor degree in Statistics from Saradha College, Salem under Madras University, Master's degree in Computer Applications from Bharathiar University, Coimbatore, Master's degree in Software Systems from BITS, Pilani, and M.Phil From Manonmaniam Sundaranar University. She is currently pursuing Ph.D. under Mother Teresa Women's University, Kodaikanal. She has 18 years of Teaching Experience and 2 years of Industrial experience. She has published number papers in various national and international conferences. She is life member of ISTE. Her field of interest includes Grid Computing, Network Security and Image Processing.

M.Madheswaran received the BE Degree from Madurai Kamaraj University in 1990, ME Degree from Birla Institute of Technology, Mesra, Ranchi, India in 1992, both in Electronics and Communication Engineering. He obtained his PhD degree in Electronics Engineering from the Institute of Technology, Banaras Hindu University, Varanasi, India, in 1999. At present he is a Principal of Muthayammal Engineering College, Rasipuram, India. He has authored over Seventy five research publications in International and National Journals and Conferences. Currently he is the chairman of IEEE India Electron Devices Society Chapter. His areas of interest are theoretical modeling and simulation of high-speed semiconductor devices for integrated optoelectronics application, Bio-optics and Bio-signal Processing. He was awarded the Young Scientist Fellowship (YSF) by the State Council for Science and Technology, TamilNadu, in 1994 and Senior Research Fellowship (SRF) by the Council of Scientific and Industrial Research (CSIR), Government of India in 1996. Also he has received YSF from SERC, Department of Science and Technology, Govt. of India. He is named in Marquis Who's Who in Science and engineering in the year 2006. He is a Member of Institute of Electrical and Electronics Engineers, Fellow of Institution of Electronics and Telecommunication Engineers, Member of Indian Society for Technical Education and Member of Institution of Engineers.

AUTHORS PROFILE

Low Power and Area Consumption Custom Networks-On-Chip Architectures Using RST Algorithms

¹P.Ezhumali ²Dr.C.Arun

¹Professor, Dept of Computer Science Engineering

²Asst. Professor, Dept of Electronics and Communication
Raalakshmi Engineering College, Thandalam-602 105, Chennai, India

Abstract: Network-on-Chip (NoC) architectures with optimized topologies have been shown to be superior to regular architectures (such as mesh) for application specific multiprocessor System-on-Chip (MPSoC) devices. The application specific NoC design problem takes, as input the system-level floorplan of the computation architecture. The objective is to generate an area and power optimized NoC topology. In this work, we consider the problem of synthesizing custom networks-on-chip (NoC) architectures that are optimized. Both the physical links and routers determine the power consumption of the NoC architecture. Our problem formulation is based on the decomposition of the problem into the inter-related steps of finding good flow partitions, and providing an optimized network implementation for the derived topologies. We used Rectilinear-Steiner-Tree (RST)-based algorithms for generating efficient and optimized network topologies. Experimental results on a variety of NoC benchmarks showed that our synthesis results were achieve reduction in power consumption and average hop count over different mesh implementations. We analyze the quality of the results and solution times of the proposed techniques by extensive experimentation with realistic benchmarks and comparisons with regular mesh-based NoC architectures.

Index Terms—Multicast routing, network-on-chip (NoC), synthesis, system-on-chip (SoC), topology.

1.Introduction

Network-on-Chip (NoC) is an emerging

paradigm for communications within large VLSI systems implemented on a single silicon chip. The layered-stack approach to the design of the on-chip intercore communications is the Network-on-Chip (NOC) methodology. In a NoC system, modules such as processor cores, memories and specialized IP blocks exchange data using a network as a "public transportation" sub-system for the information traffic. A NoC is constructed from multiple point-to-point data links interconnected by switches (a.k.a. routers), such that messages can be relayed from any source module to any destination module over several links, by making routing decisions at the switches.

A NoC is similar to a modern telecommunications network, using digital bit-packet switching over multiplexed links. Although packet switching is sometimes claimed as necessity for a NoC, there are several NoC proposals utilizing circuit-switching techniques. This definition based on routers is usually interpreted so that a single shared bus, a single crossbar switch or a point-to-point network is not NoCs but practically all other topologies are. This is somewhat confusing since all above-mentioned are networks (they enable communication between two or more devices) but they are not considered as network-on-chips. Note that some erroneously use NoC as a synonym for mesh topology although NoC paradigm does not dictate the topology. Likewise, the regularity of topology is sometimes considered as a requirement, which is, obviously, not the case in research concentrating on "application-specific NoC topology synthesis".

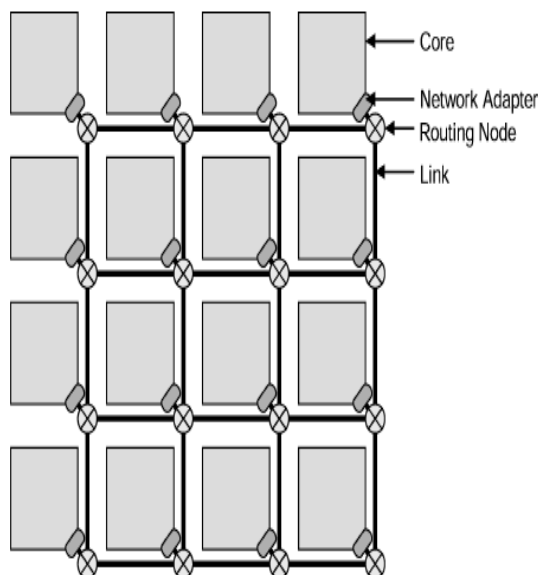


figure.1 Topological illustration of a 4-by-4 grid structured NoC.

The wires in the links of the NoC are shared by many signals. A high level of parallelism is achieved, because all links in the NoC can operate simultaneously on different data packets. Therefore, as the complexity of integrated systems keeps growing, a NoC provides enhanced performance (such as throughput) and scalability in comparison with previous communication architectures (e.g., dedicated point-to-point signal wires, shared buses, or segmented buses with bridges). Of course, the algorithms must be designed in such a way that they offer large parallelism and can hence utilize the potential of NoC.

Traditionally, ICs have been designed with dedicated point-to-point connections, with one wire dedicated to each signal. For large designs, in particular, this has several limitations from a physical design viewpoint. The wires occupy much of the area of the chip, and in nanometer CMOS technology, interconnects dominate both performance and dynamic power dissipation, as signal propagation in wires across the chip requires multiple clock cycles. NoC links can reduce

the complexity of designing wires for predictable speed, power, noise, reliability, etc., because of their regular, well-controlled structure. From a system design viewpoint, with the advent of multi-core processor systems, a network is a natural architectural choice. A NoC can provide separation between computation and communication; support modularity and IP reuse via standard interfaces, handle synchronization issues, serve as a platform for system test, and, hence, increase engineering productivity.

Although NoCs can borrow concepts and techniques from the well-established domain of computer networking, it is impractical to blindly reuse features of "classical" computer networks and symmetric multiprocessors. In particular, NoC switches should be small, energy-efficient, and fast. Neglecting these aspects along with proper, quantitative comparison was typical for early NoC research but nowadays they are considered in more detail. The routing algorithms should be implemented by simple logic, and the number of data buffers should be minimal. Network topology and properties may be application-specific. Research on NoC is now expanding very rapidly, and there are several companies and universities that are involved. Figure 1 shows how a NoC, in comparison with shared buses, could be occupied with various components as resources

2.EXISTING RELATED WORKS

So far, the communication problems faced by System on chip were tackled by making use of regular Network on chip architectures. The following are the list of popular regular NoC architectures:

- Mesh Architecture.
- Torus Architecture.
- Butterfly Fat Tree Architecture.

Extended Butterfly Fat Tree Architecture

The NoC design problem has received considerable attention in the literature. Towles and Dally [1] and Benini and De Micheli [2] motivated the NoC paradigm. Several existing NoC solutions have addressed the mapping problem to a regular mesh-based NoC architecture [3], [4]. Hu and Marculescu [3] proposed a branch-and-bound algorithm for the mapping of computation cores on to mesh-based NoC architectures. Murali *et al.* [4] described a fast algorithm for mesh-based NoC architectures that considers different routing functions, delay constraints, and bandwidth requirements. On the problem of designing custom NoC architectures without assuming existing network architecture, a number of techniques have been proposed [5]–[10]. Pinto *et al.* [7] presented techniques for the constraint-driven communication architecture synthesis of point-to-point links by using heuristic-based -way merging. Their technique is limited to topologies with specific structures that have only two routers between each source and sink pair. Ogras *et al.* [5], [6] proposed graph decomposition and long link insertion techniques for application-specific NoC architectures. Srinivasan *et al.* [8], [9] presented NoC synthesis algorithms that consider system-level floor planning, but their solutions only considered solutions based on a slicing floorplan where router locations are restricted to corners of cores and links run around cores. Murali *et al.* [10] presented an innovative deadlock-free NoC synthesis flow with detailed backend integration that also considers the floorplanning process. The proposed approach is based on the min-cut partitioning of cores to routers. This work presents a synthesis approach based on a set partitioning formulation that considers multicast traffic. Although different in topology and some other aspects, all the above papers essentially advocate the advantages of using NoCs and regularity as effective means

to design high performance SoCs. While these papers mostly focus on the concept of regular NoC architecture (discussing the overall advantages and challenges), to the best of our knowledge, our work is better than previous custom NoC synthesis formulations and efficient way to solve it.

PROPOSED SYSTEM

3.1 PROBLEM DEFINITION

- We consider the problem of synthesizing custom networks-on-chip (NoC) architectures that are optimized for a given application.
- We divide the problem statement into the following interrelated steps:

Physical topology Construction.
Power and Area Comparisons

3.2 SYSTEM ARCHITECTURE

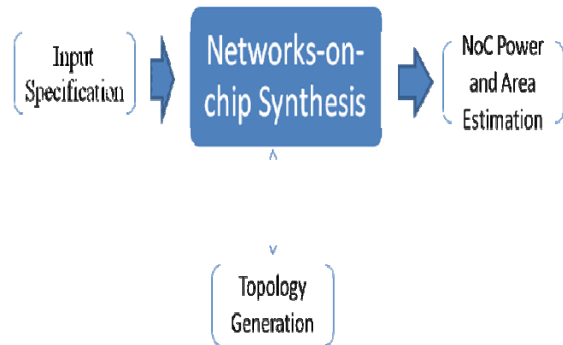


Figure. 2 Proposed System Architecture

Our NoC synthesis design flow is depicted in Figure 2. The major elements in the design flow are elaborated as follows.

Input Specification: The input specification to our design flow consists of a list of modules. As observed in recent trends, many modern SoC designs combine both hard and soft modules as well as both packet-based network communications and conventional

wiring. Modules can correspond to a variety of different types of intellectual property (IP) cores such as embedded microprocessors, large embedded memories, digital signal processors, graphics and multimedia processors, and security encryption engines, as well as custom hardware modules. These modules can come in a variety of sizes and can be either hard or soft macros, possibly as just black boxes with area and power estimates and constraints on aspect ratios. To facilitate modularity and interoperability of IP cores, packet-based communication with standard network interfaces is rapidly gaining adoption. Custom NoC architectures are being advocated as a scalable solution to packet-based communication. In general, a mixture of network-based communications and conventional wiring may be utilized as appropriate, and not all inter-module communications are necessarily over the on-chip network. For example, an embedded microprocessor may have dedicated connections to its instruction and data cache modules. Our design flow and input specification allow for both interconnection models. Below is an example of a communication demand graph:

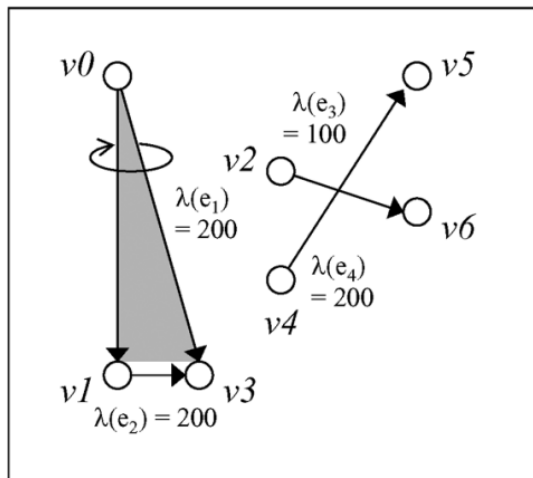


Figure 3 Sample Input Specification

NoC Synthesis: Given input specification information, the NoC synthesis step then proceeds to synthesize a NoC architecture

that is optimized for the given specification. Consider the above diagram that depicts a small illustrative example. It only shows the portion of the input specification that corresponds to the network-attached modules and their traffic flows. The nodes represent modules, edges represent traffic flows, and edge labels represent the length of the two vertices. The NoC Synthesis generates topologies based on the communication demand graph and comparing with parameters like power consumption and area usage chooses the best architecture. Below is an example of two architectures generated based on the given CDG.

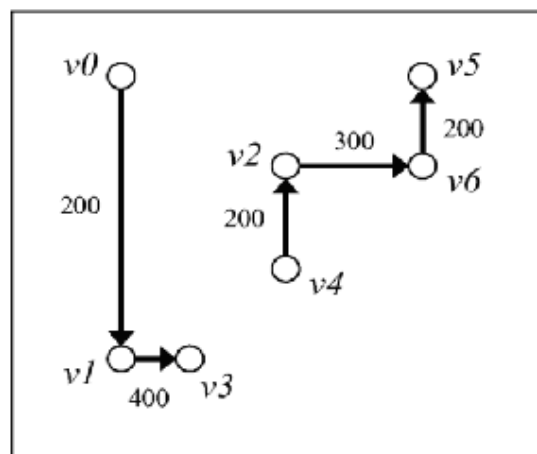
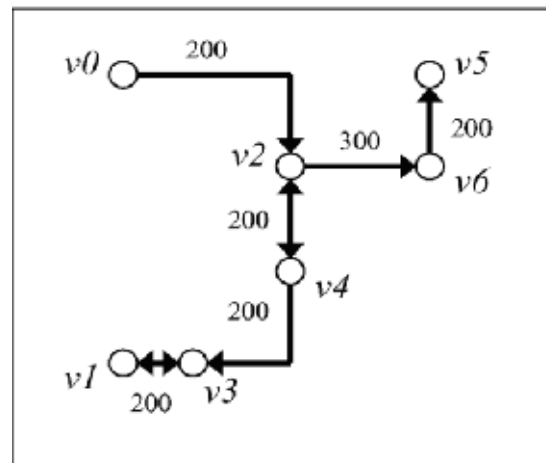


Figure 4 Sample Topologies Generated
NoC Power and Area Estimation: To evaluate the power and area of the

synthesized NoC architecture, we use a state-of-the-art NoC power-performance simulator called Orion that can provide detailed power characteristics for different power components of a router for different input/output port configurations. It accurately considers leakage power as well as dynamic switching power, which is important since it is well known that leakage power is becoming an increasingly dominating. Orion also provides area estimates based on a state-of-the-art router microarchitecture.

MODULE DESCRIPTION

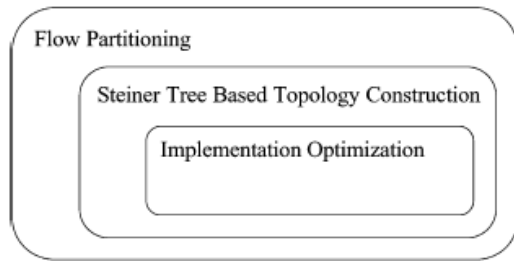


Figure 5 Formulation of Synthesis Problem

4.1 Flow Partitioning

Flow partitioning is performed in the outer loop of our synthesis formulation to explore different partitioning of flows to separate subnetworks. We make use of the following algorithm to implement flow partitioning:

4.2 STEINER TREE BASED TOPOLOGY CONSTRUCTION

For each flow partition considered, physical network topologies must be decided. In current process technologies, layout rules for implementing wires dictate physical topologies where the network links run horizontally or vertically. Thus, the problem is similar to Rectilinear Steiner Tree (RST) problem that has been extensively studied for the conventional VLSI routing problem. Given a set of nodes, the RST problem is to find a network with the

shortest edge lengths using horizontal and vertical edges such that all nodes are interconnected. The RST problem is well studied with very fast implementations available. We create an RST solver in the inner loop of flow partitioning to generate topologies for the set partitions considered.

Input: $G(V, E, \pi, \lambda)$: communication demand graph
C: specified evaluation function for implementation cost
Output: T : synthesized network architecture

```

1: initialize  $P^0 = \emptyset$ 
2: for all  $e_k \in E$ 
3:    $P^0 = P^0 \cup \{e_k\}$ 
4:    $\text{cost}(\{e_k\}) = \text{EvaluateCost}(T(\{e_k\}), C)$ 
5: end for
6:  $t = 0$ 
7: while  $|P^t| > 1$  do
8:   for all  $g_u, g_v \in P^t$  do
9:      $g_{uv} = g_u \cup g_v$ 
10:     $T(g_{uv}) = \text{SolveRST}(g_{uv})$ 
11:     $\text{cost}(g_{uv}) = \text{EvaluateCost}(T(g_{uv}), C)$ 
12:     $\beta(g_u, g_v) = \text{cost}(g_{uv}) + \sum_{g_i \in P^t, g_i \neq g_u, g_v} \text{cost}(g_i)$ 
13:  end for
14:   $(u, v) = \arg \min_{g_u, g_v \in P^t} \beta(g_u, g_v)$ 
15:   $P^{t+1} = P^t \setminus \{g_u, g_v\}$ 
16:   $P^{t+1} = P^{t+1} \cup \{g_u \cup g_v\}$ 
17:   $t = t + 1$ 
18: end while
19: for all  $t \in [0, n-1]$  do
20:    $c(P^t) = \sum_{g_u \in P^t} \text{cost}(g_u)$ 
21:    $\text{soln}[P^t] = \bigcup_{g_u \in P^t} T(g_u)$ 
22: end for
23:  $t = \arg \min_{t \in [0, n-1]} c(P^t)$ 
24:  $T = \text{soln}[P^t]$ 
25: return  $T$ 

```

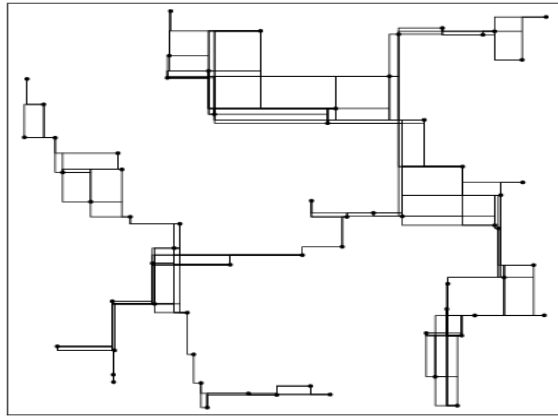
Figure 6 Flow Partitioning Algorithm

IMPLEMENTATION RESULTS

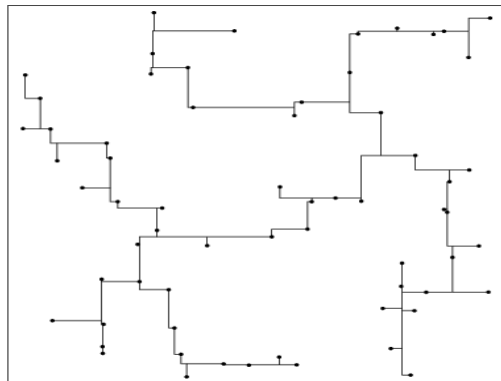
5.1. EXPERIMENTAL SETUP

We have implemented our proposed algorithm in C. In our implementation, we have designed a Rectilinear Steiner Tree solver to generate the physical network topologies in the inner loop of the algorithm. Simulator ORION 2.0 does the power and area estimates. The Results obtained are shown in a line chart for mere comparisons. A snapshot of the all the results have been shown later in this chapter. All experimental results were obtained on a 3.06-GHz Intel P4 processor machine with 512 MB of memory running Linux.

5.2. EXPERIMENTAL RESULTS



ALL FSTs: 64 Points
Figure 7 Snapshot of ALL The FSTs
Generated



Steiner Minimal Tree: 64 Points, length =
56729

Figure 8 Steiner Minimal Tree Generated

Method of Evaluation: In all our experiments, we aim to evaluate the performance of the proposed algorithms. On all benchmarks with the objective of minimizing the total area as well as power consumption of the synthesized NoC architectures. The total area as well as power consumption includes all network components. We applied the design parameters of 1 GHz clock frequency, 4-flit buffers, and 128-bit flits. For evaluation, fair direct comparison with previously published NoC

synthesis results is difficult in part because of vast differences in the parameters assumed. To evaluate the effectiveness of our algorithms, we have the full mesh implementation for each benchmark for comparison from previous published papers have been taken. These comparisons are signified to show the benefits of custom NoC architectures.

Table 1. NOC Power Comparisons

S.No	Vertices	Custom Power	MeshPower	Opt. MeshPower
1	6	0.0416	0.0990	0.0430
2	7	0.0432	0.1000	0.0500
3	8	0.0494	0.1780	0.0600
4	11	0.0617	0.2570	0.1220
5	12	0.0663	0.2720	0.1520
6	14	0.0848	0.3760	0.1310
7	20	0.0987	0.4950	0.1540
8	24	0.1034	0.6330	0.2020
9	25	0.1034	0.6420	0.2600
10	36	0.1203	1.0080	0.3000
11	44	0.1358	1.4250	0.3440

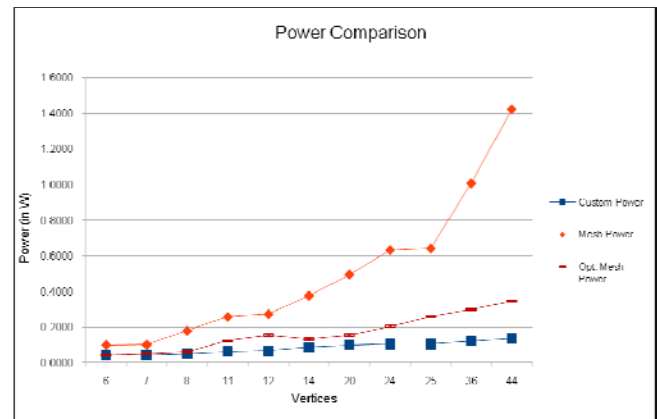


Figure 9 NoC Power Comparisons

The area results, power results, the execution times, and area as well as power improvements of that algorithm are reported. The results show the algorithm can efficiently synthesize NoC architectures that minimize power and area consumption as compared with regular topologies such as mesh and optimized mesh topologies.

Table 2. NoC Area Comparisons

S.No	Vertices	Custom Area	Opt. Mesh Area
1	6	0.1543	0.31
2	7	0.1557	0.43
3	8	0.1810	0.41
4	12	0.2252	0.48
5	14	0.3060	0.91
6	20	0.3563	1.03
7	24	0.3732	1.19
8	25	0.3753	1.81
9	36	0.4382	1.81
10	44	0.4936	2.01

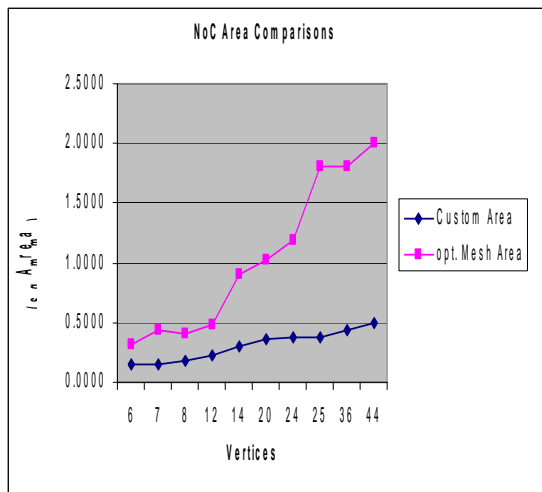


Figure 10. NoC Area Estimates

Thus, the above two line charts in figure 9 and 10 clearly show a reduction in power and area estimates of custom NoC with mesh and optimized mesh topologies. Mesh topologies was explained in chapter 2. Eliminating router ports and links that are not used forms optimized mesh topologies. The power reduction is at an average of 83.43 percent and 50 percent as compared to mesh and optimized mesh topologies respectively. The area reduction is at an average of 70.95 percent as compared to optimized mesh topologies.

6.CONCLUSION AND FUTURE WORK

In this research Works have been carried out in context related to Regular topologies like mesh, torus and etc. This work presented an idea on building customizing network on chip with the better flow partitioning and also considered power and area reduction as compared to the already presented Regular topologies, we proposed a formulation of the custom NoC synthesis problem based on the decomposition of the problem into the inter-related steps of deriving a good physical network topology, and providing an comparison in terms of area and power with the well established regular topologies. We used the algorithm called CLUSTER for systematically examining different possible set partitioning of flows, and we proposed the use of RST algorithms for constructing good physical network topologies. Our solution framework enables the decoupling of the evaluation cost function from the exploration process, thereby enabling different user objectives and constraints to be considered. Although we use Steiner trees to generate a physical network topology for each group in the set partition, the final NoC architecture synthesized is not necessarily limited to just trees as Steiner tree implementations of different groups may be connected to each other to form non-tree structures.

This work does not differentiate the routers/switches (communication modules) with the operating modules present in the chip. In near future, the work of identifying the best placement of routers and minimizing the number of routers and also the effectiveness of the customized Network on Chip in terms of other parameters like throughput, latency. Link utilization and buffer utilization can be taken into account.

REFERENCES

- [1] Shan Yan, Bill Lin, “Custom Networks-on-Chip Architectures With Multicast Routing,” *IEEE transactions on very large scale integration (VLSI) systems*, vol. 17, no. 3, march 2009.
- [2] K. Srinivasan, K. S. Chatha, and G. Konjevod, “Linear-programming based techniques for synthesis of network-on-chip architectures,” *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 14, no. 4, pp. 407–420, Apr. 2006.
- [3] K. Srinivasan, K. S. Chatha, and G. Konjevod, “Application specific network-on-chip design with guaranteed quality approximation algorithms,” in *Proc. ASPDAC*, 2007, pp. 184–190.
- [4] S. Murali, P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. De Micheli, and L. Raffo, “Designing application-specific networks on chips with floor plan information,” in *Proc. ICCAD*, 2006, pp. 355–362.
- [5] L. Zhang, H. Chen, H. Chen, B. Yao, K. Hamilton, and C.-K. Cheng, “Repeated on-chip interconnect analysis and evaluation of delay, power, and bandwidth metrics under different design goals,” in *Proc. ISQED*, 2007, pp. 251–256.
- [6] R. Mullins, “Minimizing dynamic power consumption in on-chip networks,” in *Proc. Int. Symp. Syst.-on-Chip*, 2006, pp. 1–4.
- [7] C. -W. Lin, S. -Y. Chen, C. -F. Li, Y. -W. Chang, and C. -L. Yang, “Efficient obstacle-avoiding rectilinear Steiner tree construction,” in *Proc.Int. Symp. Phys. Des.* 2007, pp. 127–134.
- [8] D. Greenfield, A. Banerjee, J. -G. Lee, and S. Moore, “Implications of rent’s rule for NoC design and its fault-tolerance,” in *Proc. NOCS*, May 2007, pp. 283–294.
- [9] S. Yan and B. Lin, “Application-specific network-on-chip architecture synthesis based on set partitions and Steiner trees,” in *Proc. ASPDAC*, 2008, pp. 277–282.
- [10] Xilinx, San Jose, CA, “UMC delivers leading-edge 65 nm FPGAs toXilinx,” *Des. Reuse*, Nov. 8, 2006 [Online]. Available: <http://www.design-reuse.com/news/14644/umc-edge-65nm-fpgas-xilinx.html>
- [11] P. Gratz, K. Sankaralingam, H. Hanson, P. Shivakumar, R. McDonald, S. W. Keckler, and D. Burger, “Implementation and evaluation of a dynamically routed processor operand network,” in *Proc. NOCS*, May 2007, pp. 7–17.
- [12] N. Enright-Jerger, M. Lipasti, and L.-S. Peh, “Circuit-switched coherence,” *IEEE Computer. Arch. Lett.* vol. 6, no. 1, pp. 193–202, Mar. 2007.
- [13]. Shan Yan, Student Member, IEEE, and Bill Lin, Senior Member, IEEE “Custom Networks-on-Chip Architectures With Multicast Routing” *IEEE Transactions On Very Large Scale Integration (VLSI) Systems*, Vol. 17, No. 3, Pp 342-355, March 2009.



Ezhumalai Periyathambi received the B.E degree in Computer Science and engineering from Madras University, Chennai , India in 1992 and Master Technology (M.Tech.,) in computer science and Engineering from J N T University, Hyderabad, India in 2006. He is currently

working towards the Ph.D degree in Department of Information and Communication, Anna University, Chennai, India. He is working as Professor in the Department of Computer Science and Engineering , Rajalakshmi Engineering College, Chennai, Tamilnadu, India. His research in reconfigurable architecture, Multi-Core Technology CAD – Algorithms for VLSI Architecture. Theoretical Computer Science. And mobile computing.

Prediction of Epileptic form Activity in Brain Electroencephalogram Waves using Support vector machine

¹Pavithra Devi S T

M.Phil Research Scholar
PSGR Krishnammal College for Women
Coimbatore Tamilnadu, INDIA

²Vijaya M S

Assistant Professor and Head
GRG School of Applied Computer
Technology
PSGR Krishnammal College for Women
Coimbatore Tamilnadu, INDIA

ABSTRACT

Human brain is a highly complex structure composed of millions of nerve cells. Their activity is usually well organized with mechanisms for self-regulation. The neurons are responsible for a range of functions, including consciousness and bodily functions and postures. A sudden temporary interruption in some or all of these functions is called a seizure. Epilepsy is a brain disorder that causes people to have recurring seizures. Electroencephalogram (EEG) is an important diagnostic test for diagnosing epilepsy because it records the electrical activity of the brain. This paper investigates the modeling of epilepsy prediction using Support Vector Machine, a supervised learning algorithm. The prediction model has been employed by training support vector machine with evocative features derived from EEG data of 324 patients and from the experimental results it is observed that the SVM model with RBF kernel produces 86% of accuracy in predicting epilepsy in human brain.

Keywords

Support Vector Machine, Epilepsy, Prediction, Supervised Learning.

1. INTRODUCTION

Epilepsy is a disorder characterized by recurrent seizures of cerebral origin, presenting with episodes of sensory, motor or autonomic phenomenon with or without loss of consciousness. Epilepsy is a disorder of the central nervous system, specifically the brain [1]. Brain is one of the most vital organs of humans, controlling the coordination of human muscles and nerves. Epileptic seizures typically lead to an assortment of temporal changes in perception and behavior. Based on the physiological characteristics of epilepsy and the abnormality in the brain, the kind of epilepsy is determined. Epilepsy is broadly classified into absence epilepsy, simple partial, complex partial and general epilepsy. Absence epilepsy is a brief episode of staring. It usually begins between ages 4 and 14. It may also continue to adolescence or even adulthood. Simple partial epilepsy affects only a small region of the brain, often the hippocampus. It can also include sensory disturbances, such as smelling or

hearing things that are not there, or having a sudden flood of emotions. Complex partial epilepsy usually starts in a small area of the temporal lobe or frontal lobe of the brain. In general epilepsy the patient becomes unconscious the patient has a general tonic contraction of all their muscles, followed by alternating clonic contractions. It affects the entire brain.

Various diagnostic techniques like Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Electroencephalogram (EEG), and Positron Emission Tomography (PET) are commonly presented. Electroencephalography (EEG) is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a short period of time, usually 20–40 minutes, as recorded from multiple electrodes placed on the scalp. The Electroencephalograph (EEG) signal is one of the most widely signal used in the bioinformatics field due to its rich information about human tasks for epilepsy identification because of its characteristics like frequency range, spatial distributions and peak frequency. EEG waves are observed by neurologists based on spectra waveform of the signal to identify the presence of epilepsy.

Machine learning provides methods, techniques and tools, which help to learn automatically and to make accurate predictions based on past observations. Current empirical results prove that machine learning approach is well-matched for analyzing medical data and machine learning techniques produce promising research results to medical domains.

Forrest Sheng Bao carried out the work and developed a neural network based model for Epilepsy diagnosis using EEG [1]. Piotr Mirowski carried out the work and implemented a model based on classification of patterns of EEG synchronization for seizure prediction using neural network [2]. Suleiman A.B. R. proposed a new approach for describing and classifying the EEG brain natural oscillations such as delta, theta, alpha, and beta frequencies

using Wigner-Ville analysis with Choi-Willians filtering and neural network [3].

The motivation behind the research reported in this paper is to predict the presence of epilepsy in human brain. Supervised learning technique, a kind of machine learning algorithm is used to model the epilepsy prediction problem as classification task to assist physician for accurate prediction of epilepsy in patients.

In this paper, the prospective benefits of supervised learning algorithm namely support vector machine are made use of for the computerized prediction of epilepsy. The proposed SVM based epilepsy prediction model is shown in Figure 1.

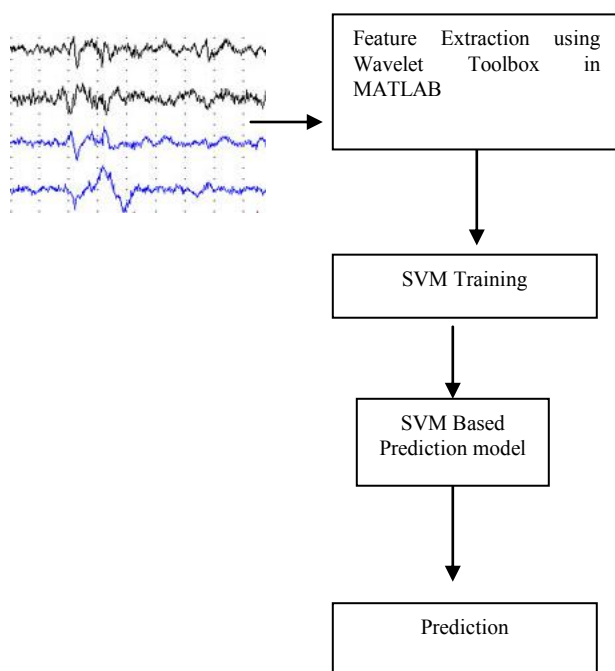


Figure 1. Proposed SVM based epilepsy prediction model

2. DATA ACQUISITION

EEGs show continuous oscillating electric activity. The amplitude and the patterns are determined by the overall excitation of the brain which in turn depends on the activity of the reticular activating system in the brain stem. Amplitudes on the surface of the brain can be up to 10 mV, those on the surface of the scalp range up to 100 mV. Frequencies range from 0.5 to 100 Hz. The pattern changes markedly between states of sleep and wakefulness. Distinct patterns are seen in epilepsy and five classes of wave groups are described as alpha, beta, gamma, delta and theta.

- Alpha waves contain frequencies between 8 and 13 Hz with amplitude less than 10 mV. It found in normal people who are awake and resting quietly, not being engaged in intense

mental activity. Their amplitude is highest in the occipital region. When the person is asleep, the alpha waves disappear. When the person is alert and their attention is directed to a specific activity, the alpha waves are replaced by asynchronous waves of higher frequency and lower amplitude.

- Beta waves have a frequency range of 14 to 22 Hz, extending to 50 Hz under intense mental activity. It has their maximum amplitude (less than 20 mV) on the parietal and frontal regions of the scalp. There are two types: beta I waves, lower frequencies which disappear during mental activity, and beta II waves, higher frequencies which appear during tension and intense mental activity.

- Gamma waves have frequencies between 22 and 30 Hz with amplitude of less than 2 mV peak-to-peak and are found when the subject is paying attention or is having some other sensory stimulation.

- Theta waves have a frequency range between 4 to 7 Hz with amplitude of less than 100 mV. It occurs mainly in the parietal and temporal regions in sleep and also in children when awake, and during emotional stress in some adults, particularly during disappointment and frustration. Sudden removal of something causing pleasure will cause about 20 s of theta waves.

- Delta waves have frequency content between 0.5 and 4 Hz with an amplitude less than 100 mV. It occurs during deep sleep, during infancy and in serious organic brain disease. They will occur after transactions of the upper brain stem separating the reticular activating system from the cerebral cortex. They are found in the central cerebrum, mostly the parietal lobes.

Five sets of images namely Normal Epilepsy, Absence Epilepsy, Simple Partial Epilepsy, Complex Partial Epilepsy and General Epilepsy are taken into consideration.

3. FEATURE EXTRACTION

Feature extraction process plays a very important role on the classification. Fourier transformation method, discrete transformation method and continuous transformation methods are normally available to extract features that characterize EEG signals. The wavelet transform (WT) provides very general techniques which can be applied to many tasks in signal processing. Wavelets are ideally suited for the analysis of sudden short-duration signal changes.

In the proposed model, EEG signal analysis and feature extraction have been performed using Discrete Wavelet Transform (DWT). The DWT is a extraordinary case of the WT that provides a compact representation of a signal in time and frequency that can be computed efficiently.

The DWT is defined by the following equation:

$$\Psi_{(a,b)}(t) = 2^{-a/2} \psi(2^{-a/2}(t-b)) \quad (1)$$

where a is a scales and b is positions of the wavelet mother $\psi(t)$ is a time function with finite energy. Choosing scales and positions are based on powers of two, which are called dyadic scales and positions ($a = 2^{-j}$; $b_{j,k} = 2^j k$) (j and k integers). Equation (1) shows that it is possible to build a wavelet for any function by dilating a function $\psi(t)$ with a coefficient 2^j , and translating the resulting function on a grid whose interval is proportional to 2^{-j} .

The selection of appropriate wavelet and the number of decomposition levels is very important in analysis of signals using the WT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients. The smoothing feature of the Daubechies wavelet of order 2 (db2) made it more suitable to detect changes of the signals. Thus, the wavelet coefficients are computed using db2. The frequency bands corresponding to different levels of decomposition for db2 with a sampling frequency of 256 Hz. The discrete wavelet coefficients are computed using the MATLAB wavelet toolbox.

The purpose of feature extraction is to reduce the size of the original dataset by measuring certain properties or features that distinguish one input pattern from another. The various measurements based on statistical features from EEG are extracted. The extracted features provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the signals into a feature space.

The statistical feature of the wavelet coefficients in each subband such as energy, entropy, Minimum subband, maximum subband, mean, and standard deviation are used to investigate the adequacy for the discrimination of normal and abnormal patients. The following statistical features have been derived using the following.

Entropy is the diminished capacity for spontaneous changes in signals.

$$\text{Entropy} = \sum_{i,j} P(i,j) \log P(i,j) \quad (2)$$

Where $P(i,j)$ reflects the distribution of the probability of occurrence of each signal (i, j are integer).

Energy – Provides the sum of squared elements in the wavelet. This is also known as uniformity or the angular second moment.

The energy is computed using E is given by

$$E = \sum_{i=1}^n x_i^2 / N \quad (3)$$

where x_i is signal value, values are present in waves is denoted as n . Total number of signal is N

Maximum Subband – It generate maximum of the wavelet coefficients in each subband is calculated using

$$M_{ax} = \text{Max}(x_i) \quad (4)$$

where $\text{max}(x_i)$ is maximum number of signal value.

Mean – It is defined as average value of a distribution of the wavelet coefficients in each subband which is given by

$$E = \sum_{i=1}^n x_i / N \quad (5)$$

where x_i is signal and total number of signal is present in the wavelet is N

Minimum Subband – calculate minimum of the wavelet coefficients in each subband is defined as

$$M_{in} = \text{Min}(x_i) \quad (6)$$

Where $\text{min}(x_i)$ is minimum number of signal value.

Standard deviation - standard deviation of each subband is defined as σ . This feature provide information about the amount of change of the frequency distribution.

$$\sigma = \sum_{i=1}^n (x_i - \mu)^2 \quad (7)$$

where \sum is sum of squared elements in the wavelet, x is signal value and μ is a mean of the corresponding signal(x_i).

Thus a total of 21 statistical feature are extracted from EEG signal for each subband for preparing dataset.

4. SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a kind of learning machine based on statistical learning theory. SVM is basically applied to model pattern classification task. SVM first, maps the input vectors into feature vectors in feature space with a higher dimension, either linearly or non-linearly. Then, within the feature space SVM constructs a hyperplane which separates two classes. SVM training always seeks a global optimized solution and avoids over-fitting, thus it has the ability to deal with a large number of features. The machine is presented with a set of training examples, (x_i, y_i) where the x_i is the real world data instances and the y_i are the labels indicating which class the instance belongs to. For the two class pattern recognition problem, $y_i = +1$ or $y_i = -1$. A training example (x_i, y_i) is called positive if $y_i = +1$ and negative otherwise. SVMs construct a hyperplane that separates two classes and tries to achieve maximum separation between the classes. Separating the classes with a large margin minimizes a bound on the expected generalization error.

The simplest model of SVM called Maximal Margin classifier, constructs a linear separator (an optimal hyperplane) given by $w^T x - y = 0$ between two classes of examples. The free parameters are a vector of weights w which is orthogonal to the hyper plane and a threshold value. These parameters are obtained by solving the following optimization problem using Lagrangian duality.

$$\text{Minimize} = \frac{1}{2} \|w\|^2$$

$$\text{Subject to } D_{ii} (w^T x_i - \gamma) \geq 1, i = 1, \dots, l. \quad (8)$$

where D_{ii} corresponds to class labels $+1$ and -1 . The instances with non-null weights are called support vectors. In the presence of outliers and wrongly classified training examples it may be useful to allow some training errors in order to avoid over fitting. A vector of slack variables ξ_i that measure the amount of violation of the constraints is introduced and the optimization problem referred to as soft margin is given below. In this formulation the contribution to the objective function of margin maximization and training errors can be balanced through the use of regularization parameter C .

The following decision rule is used to correctly predict the class of new instance with a minimum error.

$$f(x) = \text{sgn}[w^T x - \gamma]$$

The advantage of the dual formulation is that it permits an efficient learning of non-linear SVM separators, by introducing kernel functions. Technically, a kernel function calculates a dot product between two vectors that have been (non-linearly) mapped into a high dimensional feature space. Since there is no need to perform this mapping explicitly, the training is still feasible although the

dimension of the real feature space can be very high or even infinite. The parameters are obtained by solving the following non linear SVM formulation (in Matrix form),

$$\text{Minimize } L_D(u) = \frac{1}{2} u^T Q u - e^T u \quad (9)$$

$$d^T u = 0 \quad 0 \leq u \leq C e$$

where d and K - the Kernel Matrix. $Q = DKD$.

The Kernel Function K (AAT) (polynomial or Gaussian) is used to construct hyperplane in the feature space, which separates two classes linearly, by performing computations in the input space.

$$f(x) = \text{sgn}(K(x, x_i^T) * u - \gamma)$$

Where u - the Lagrangian multipliers. In general the larger the margin the lower the generalization error of the classifier.

5. EXPERIMENTAL SETUP

The data investigation and epilepsy prediction is carried out using SVMlight¹ for machine learning. Five categories of feature vectors are labeled as 1 for Absence, 2 for General, 3 for Complex Partial Epilepsy, 4 for Normal Epilepsy and 5 for Simple Partial Epilepsy. The training dataset used for epilepsy prediction modeling consists of about 324 images, where each category consists of about 65.

The dataset has been trained using SVM with linear, polynomial and RBF kernel and with different parameter settings for d , γ and C -regularization parameter. The parameters d and γ are related with polynomial kernel and RBF kernel respectively.

The 10 fold cross validation method is used for evaluating the performance of the SVM based trained models. The performance of the models is evaluated based on prediction accuracy of the models and learning time.

6. RESULTS AND DISCUSSION

The cross validation outcome of the trained models based on support vector machine with linear kernel is shown Table I.

Table I. SVM Linear kernel

Linear SVM	C=0.1	C=0.2	C=0.3	C=0.4
Accuracy (%)	70	72	76	78
Time(secs)	0.01	0.02	0.02	0.03

¹ SVMlight is an open source tool.

http://www.cs.cornell.edu/people/tj/svm_light/

The outcome of the model based on SVM with polynomial kernel and with parameters d and C are shown in Table II.

Table II. SVM Polynomial kernel

d	$C=0.1$		$C=0.2$		$C=0.3$		$C=0.4$	
	1	2	1	2	1	2	1	2
Accuracy (%)	70	80	82	80	80	81	74	75
Time(secs)	0.2	0.1	0.2	0.6	0.3	0.1	0.3	0.4

The predictive accuracy of the non-linear support vector machine with the parameter gamma (g) of RBF kernel and the regularization parameter C is shown in Table III.

Table III. SVM RBF kernel

g	$C=0.1$		$C=0.2$		$C=0.3$		$C=0.4$	
	1	2	1	2	1	2	1	2
Accuracy (%)	80	83	83	81	83	86	85	77
Time(secs)	0.2	0.3	0.4	0.4	0.5	1.5	1.6	1.2

The average and comparative performance of the SVM based prediction model in terms of predictive accuracy and learning time is given in Table IV and shown in Figure 1 and Figure 2.

Table IV. Overall performance of three models

Kernel type	Accuracy	Learning time
Linear	84.96%	0.027 secs
Polynomial	90.12%	0.362 secs
RBF	93.87%	0.787 secs

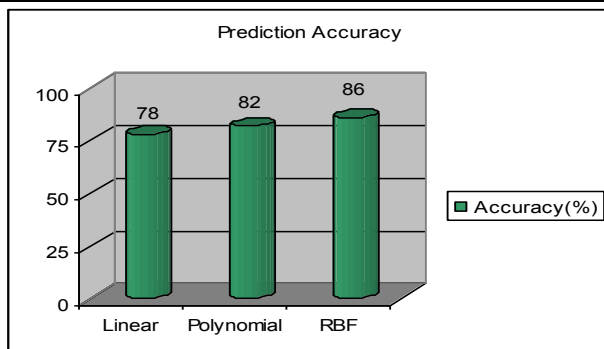


Figure 2. Prediction Accuracy

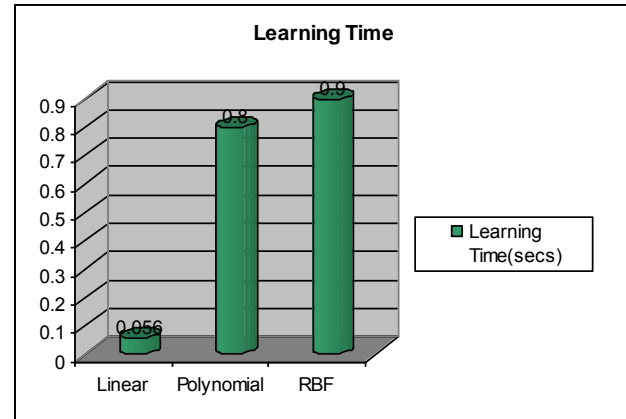


Figure 3: Learning Time

As far as the epilepsy predictions task is anxious, accuracy plays major role in determining the performance of the epilepsy trained model than considering the learning time. From the above results, it is found that the predictive accuracy shown by SVM with RBF kernel with parameters $C=0.2$ and $g=2$ is higher than the SVM with linear and polynomial kernel.

7. CONCLUSION

This paper elucidates the modeling of the epileptic seizure prediction task as multi-class classification problem and the implementation of supervised learning algorithm, support vector machine. The performance of SVM based epilepsy prediction models is evaluated using 10 fold cross validation and the results are analyzed. The results indicate that the support vector machine with RBF kernel provide the high prediction accuracy compared to other kernels. SVM is better than conventional methods and show good performance in all experiments it is very flexible and more powerful because of its robustness. It is hoped that more interesting results will follow on further exploration of data.

8. ACKNOWLEDGMENT

The author would like to thank the Management and Hospital, Coimbatore for providing the EEG data.

9. REFERENCES

- [1] Forrest Sheng Bao, Jue-Ming Gao, Jing Hu, Donald Y. C. Lie, Yuanlin Zhang, and K. J. Oommen. "Automated Epilepsy Diagnosis Using Interictal Scalp EEG". 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.
- [2] Piotr Mirowski MSc*, Deepak Madhavan Yann Le Cun, uben Kuzniecky "Classification of Patterns of EEG Synchronization for Seizure Prediction".
- [3] A. R.Sulaiman, "Joint Time - Frequency Analysis and Its application for Non - Stationary Signals", Ph.D. Thesis Elect. Eng. Dept., University of Mosul, 2001.
- [4] Webster, J. G., "Medical Instrumentation Application and esign", 2nd ed., New York: Wiley, 1995.

- [5] Nello Cristianini and John Shawe - Taylor. "An Introduction to Support Vector Machines and other kernel - based learning methods" Cambridge University Press, 2000.
- [6] K. Crammer and Y. Singer. "On the Algorithmic implementation of Multi - class SVMs, JMLR, 2001. Vojislav Kecman: "Learning and Soft Computing — Support Vector Machines, Neural Networks, Fuzzy Logic Systems", The MIT Press, Cambridge, MA, 2001.
- [7] Chui, C.K. (1992a), "Wavelets: a tutorial in theory and applications", Academic Press.
- [8] Ian H. Witten, Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, Sally.
- [9] Ian H. Witten, Eibe Frank. : Data Mining – Practical Machine Learning Tools and Techniques. 2nd edn. Elsevier. (2005).
- [10] Joachims T, Schölkopf B, Burges C, Smola A, "Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning", 1999, MIT Press, Cambridge, MA, USA.
- [11] John Shawe-Taylor, Nello Cristianini, "Support Vector Machines and other kernel-based learning methods", 2000, Cambridge University Press, UK.
- [12] Soman K.P, Loganathan R, Ajay V, "Machine Learning with SVM and other Kernel Methods", 2009, PHI, India.
- [13] Crammer Koby, Yoram Singer, "On the Algorithmic Implementation of Multi-class Kernel-based Vector Machines", Journal of Machine Learning Research, MIT Press, Cambridge, MA, USA, 2001, Vol.2 Page 265-292.

Deployment of Intelligent Agents in Cognitive Networks

Huda Fatima	Dr.Sateesh Kumar Pradhan	Mohiuddin Ali Khan	Dr. G.N.Dash
Dept. of CS	Dept. of Comp.Engineering	Dept. of Comp. Networks	Dept. of Comp. Science
Jazan University	King Khalid University	Jazan University	Sambalpur University
Jazan, K.S.A	Abha, K.S.A	Jazan, K.S.A	Orissa, India

Abstract—Every organization faces challenging task in the designing of the communication network in order to make its efficiency smoother by the increasing complexities. Therefore, we have to proposed a concept of cognitive network and how the intelligent agents are deployed to overcome the challenges. With the tremendous expansion of networks across the globe, the deployment of intelligent agents in cognitive networks contributes as an efficient, reliable and challenging task for the researchers. In this paper, we survey the existing research work on cognitive networks and later we provide the artificial intelligent techniques that are potentially suitable for the development of cognitive networks.

Keywords: *Artificial Intelligence, Cognitive network, Intelligent agents.*

I. INTRODUCTION:

One of the fastest growing areas is the information and communication technologies. These changes have an immediate impact on diverse aspects of the modern society, which includes inter-human relations, economy, education & entertainment. In this respect, the development of reliable, flexible and future-proof infrastructure should be capable of increasing the users' quality of life by providing services such as e-health, e-learning and e-payments. In order to meet the demand of the increased complexity, future networks should be easily

maintainable and their capabilities should be continuously improved and upgraded by relying as little as possible on human intervention. Therefore the network research community proposed a new concept of networking: The Cognitive Network. What is a Cognitive Network and how are the intelligent agents deployed is what we have presented here.

Cognitive networks

In this section, we analyze several existing definitions for cognitive networks, and we argue that two elements are essential for developing a cognitive network (CN): the knowledge representation and the cognition loop. Next, we discuss the framework proposed in [2] for introducing cognition to communication networks. The main part of the section focuses on methods from AI that seem applicable for developing CNs. We provide a summary of several types of intelligent agents (IAs), map them to the functional states of the cognitive loop. As we go along, we also refer to existing research on CNs which makes use of the respective type of IA, where available. How it started? The word cognitive refers to an entity that is able to perform some kind of conscious intellectual activity such as thinking, reasoning, learning or remembering in order to make sense of its surroundings. This word was first used in communication networks to refer to a technology by Mitola as he introduced the cognitive radio [4].

We would like to emphasize that, according to the dictionary [9], the word cognitive used as an adjective to a noun means: of, relating to, being, or involving conscious intellectual activity (as thinking, reasoning, or remembering); based on or capable of being reduced to empirical factual knowledge.

In [2], the authors define the CN as a network with a cognitive process that can perceive current network conditions, plan, decide, act on those conditions, learn from the consequences of its actions, all while following end-to-end goals. This loop, the cognition loop, senses the environment, plans actions according to input from sensors and network policies, decides which scenario fits best its end-to-end purpose using a reasoning engine, and finally acts on the chosen scenario as discussed in the previous section. The system learns from the past (situations, plans, decisions, actions) and uses this knowledge to improve the decisions in the future.

This definition of CN does not explicitly mention the knowledge of the network; it only describes the cognitive loop and adds end-to-end goals that would distinguish it from CR or so called cognitive layers [2]. We consider this definition of CN incomplete since it lacks knowledge which is an important component of a cognitive system as discussed so far in this paper and also in [2,4,6,8].

The cognitive process can operate in a centralized way, spanning over a large network, or in a totally distributed manner at a device level. In the first case, it might be too expensive to centralize all the network specific information that the cognition loop requires, while in the second case there might be too little knowledge available to pursue end-to-end network goals. In reality, the deployment of the cognitive functionality in a network will depend on the network specific problems and will be an engineering decision. However, it is important that the cognitive framework is designed in such way as to be modular, easily upgradeable and scalable in order to be able to accommodate existing as well as next generation technologies and applications.

The capabilities of a Cognitive Network can be highly distributed or extremely centralized. In general, a Cognitive network is formed of a set of

distributed cognitive entities (agents) which are somehow “smart” as they have certain reasoning capabilities to be connected to the network. The entities in such a network interact with each other, they can cooperate, act selfishly or a combination of the two. While functioning in this environment, the entities are able to learn and take decisions in such way as to reach an end-to-end goal. These end-to-end goals are dictated by the business and user requirements [2,4]. Developing and maintaining such a network is an extremely challenging task and has enormous potential, especially in the area of network management.

A Cognitive Network needs to evolve overtime: its set of technologies has to be updated by removing deprecated and adding new ones; its set of tools that help managing complexity should be added and removed in a plug and play fashion. Thus, the architecture of cognitive network should be flexible and should lead to a modular and highly scalable infrastructure. Furthermore, the cognitive network must be self aware : it should be able to determine appropriate actions to achieve goals and to learn while doing all these. It should be self-configuring, self-optimizing, self-healing and self-protecting in a cognitive way.

In this paper, we analyze some recent trends in the development of communication networks and investigate in more detail the concept of cognitive network. Cognitive networks are promising to be the major step towards efficient and automatic management of increasing complexity of communication networks.

Cyclic Process in Cognitive Network.

All systems that are able to adjust their functioning according to changes in their environment are based on feedback information. Cognitive networks are no exception in this respect, so they will also use a control loop, also called cognition cycle [7, p. 7], feedback loop [2], context based adaptation loop [8]. According to Thomas et al. [2], the loop employed by a cognitive network should be based on the concept of the Observe-Orient-Decide-Act loop originally used in the military, augmented by learning and following end-to-end goals to achieve cognition. In [8], the loop also has a communicating capability for communicating with other loops in a distributed environment.

The cognition cycle as described by Mitola [7, p. 8] features the following states: observe, orient, plan, decide, act and learn. It uses the orient module for classifying stimuli and does not explicitly encompass policies.

Cycle management:

In [10], the authors investigate a cognitive agent for wireless network selection which is designed to hide the complexity of the wireless environment from the user. The selection problem is decomposed

into four elements that enhance the agent to select the network which is most suitable to user preferences. First, user's feedback that the decision making process will be used is captured. Second, the available services are evaluated against learned user preferences. Third, the agent decides when to change services and which new service to select based on user's preferences, context and goals.

Fourth, the value of previously unseen services is predicted. Using this approach, the agent continuously monitors the wireless environment and selects the best service according to the current model of user preferences. However, when the user is unsatisfied (or changes preferences), The model is updated and a new selection is made to satisfy preferences. A Cognitive Resource Manager (CRM) and its conceptual architecture are introduced in [14]. The CRM's functioning is based on a cognition cycle adapted from Mitola [7] and aims at enabling autonomic optimization of the communication stack as a whole, thus acting as an intelligent vertical calibration(Fig.1). The intelligence would be based upon methods from the field of AI.

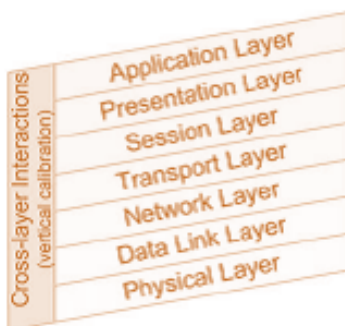


Fig 1. Open Systems Interconnection (OSI) model

Loop for security

The CycSecure application [12] makes use of an incomplete cognitive loop. It uses daemons installed on machines in the network that collect local information and send it to the server when polled. A human operator can examine and modify the network model, query and view network statistics. The system is able to generate possible attack plans based on the information gathered from the system and the internal knowledge base. Based on these attack plans, the human operator can decide for remedy measures to increase the security of the system.

Communication Requirements and research directions

In the history of telecommunications, development has always been driven by humans need to communicate, i.e. reliably transmit ever increasing amount of information across increasing distances. However, communication networks became increasingly complex and more difficult to manage, requiring increasingly specialized tools and human operators for their maintenance, configuration and optimization. From the user's point of view the necessities in the world of telecommunications, as it is today, are : higher bandwidth or alternative solutions capable of accommodating the traffic . These necessities derive from the user's thirst for digital content.

From the network operators' point of view, some of the main necessities are: complexity, management, security, scalability, fault tolerance, fast integration of new technologies and a good business model [6]. The network operator has to create adequate premises for delivering the digital content.

These user's and network operators necessities are actually forming the basis for research activities currently underway in the area of cognitive networks. In general, research directions in communications can be classified in 8 broad categories: theory, signal processing, networks, software, user satisfaction, security, management and next generation protocols and architectures. In an attempt to obtain an objective big picture of the trends in research areas as well as quantitative estimation of the ongoing work, we used ontogeny, a semi-automatic ontology editor [6] to analyze the

conference proceedings of IEEE Globecom 2006 and 2007, totaling 2011 papers

Artificial intelligence:

Artificial intelligence is concerned with intelligent behavior in artifacts. Intelligent behavior, in turn, involves perception, reasoning, learning, communicating, and acting in complex environments. Artificial Intelligence has as one of its long-term goals the development of machines that can do these things as well as humans can, or possibly even better. Another goal of AI is to understand behavior whether it occurs in machines or in humans or other animals

Intelligent agents developed in a couple of streams of work, among them is cybernetics [Wiener 198], cognitive psychology, Computational linguistics [Chomsky 1914], and adoptive control theory [Widrow & Hoff 1960], also contributed to the intellectual matrix developed by Artificial intelligence.

Intelligent Agents:

Intelligent agents in Artificial intelligence react, plan, reason and learn in an environment more or less compatible with its abilities and goals. Here we shall see how the actions of other agents can be anticipated in each agent's own planning, and indeed, how an agent can even affect the actions of other agents in the service of its own goals. To predict what another agent will do, we need methods for one agent to model another; to affect what another agent will do. There are two kinds of models used by agents, iconic and feature-based. An iconic model of the environment attempts to simulate relevant aspects of the environment; a feature-based model attempts to describe the environment—perhaps by formulas in the predicate calculus. The agents that we deploy can use either an iconic or a feature-based model of the other agents' cognitive structure. And the other agent itself might be presumed to be using either an iconic or feature-based model. The four possibilities are shown in table 1 along with the modeling strategy each one provokes.

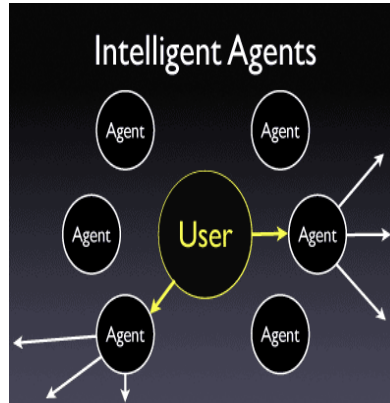


Figure 2. Intelligent agents for Cognitive Networks

The starting point towards developing a CN is the intelligent agent (IA). This section presents existing and emerging AI techniques that can prove useful for developing agents for CNs. According to Russell and Norvig [13, p. 42], an agent is central to AI. It is an entity that perceives the environment through sensors and acts upon that environment through actuators. This is the so-called “weak” definition of agency while “stronger” definitions take into account functions and characteristics of the agent [14, p. 8, 13, p. 42]. Among different classifications of agents, we will consider as a reference, the one established at IBM, which uses three dimensions to describe agents (see Fig.6). The first dimension is the Agency, which determines the degree of “autonomy and authority vested in the agent”. The second dimension is the Intelligence, which describes the degree of reasoning and learned behavior. Finally, the third dimension is Mobility, which specifies the degree to which agents travel through the network [14, p. 9]. Current networks operate via message passing (i.e. IP packets between two routers or primitives between TCP and IP) where the receiver takes an action as a consequence

of the received message. This type of operation is asynchronous and is characteristic to expert systems [14, p. 9, 15]. This approach permitted loose coupling of complex systems (e.g. communication networks). However, this approach permits the lowest degree of autonomy according to Fig. 1. On the Intelligence axis, some of the current communication systems do not even reach the lowest level as they do not even allow specification of preferences (e.g. QoS specifications). In this respect, CNs are expected to enhance the level of intelligence of current communication systems by incorporating so-called

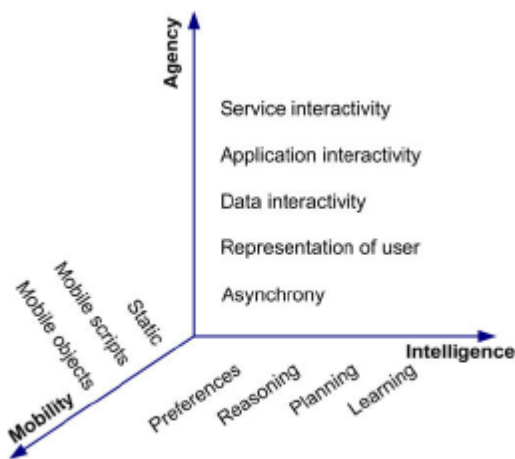
Intelligent Agents (IAs) in the KP. On the Agency axis, IAs can perform actions on behalf of the user,

more specifically they can interact with data, applications or services. On the Intelligence axis, IAs can hold a model(i.e. user, system, environment, etc.), perform reasoning, planning and learning. These actions are exactly the same as the ones desired from CN and can be found in the states of the cognition loop (see Plan, Decide, Act, Learn and Policy Fig.4).



Fig. 3. The cognition loop.

Networks of the future will make use of agents to improve their performance with respect to all three axes in Fig 1.



In the case of CNs, the main improvement is achieved with respect to the Intelligence axis. Therefore, in the remainder of the section we focus on describing utility of Intelligent Agents for these networks. We also emphasize the correspondence between Intelligent Agents and the states of the cognition loop. From the intelligence point of view,

the minimal requirement for an Intelligent Agents in general is to hold a model and be able to reason based on this model. These IAs (Intelligent Agents) are also called knowledge-based agents. Reasoning can take place upon two types of knowledge: certain (true, false and unknown) and uncertain. Reasoning under certain knowledge is accomplished by logical agents. In this respect, agents “can form representations of the world, use a process of [logical] inference to derive new representations about the world, and use these new representations to deduce what to do” [13, p. 191]. Logical agents use symbolic knowledge representations, so called artificial languages, and typically first-order logic to infer new facts. These representations also support semantic querying. Agents that have incomplete or uncertain information use decision theory and are also called decision theoretic agents. These agents use knowledge representations specific for uncertain domains (i.e. full joint distributions can constitute the knowledge base) to reason. Then they perform probabilistic inference, which is the computation of posterior probabilities from the observed evidence.

Conclusions

The recently emerging CN concept is promising to be the right answer to emerging challenges of the network management. In this paper we surveyed existing work on CNs. We first analyzed recent research trends in communications. We mapped existing AI techniques to the states of the cognition loop and identified challenges for research in AI from which CNs could benefit. We concluded the paper with identification of standardization activities related to or potentially benefiting from the research in the area of CNs.

The discussions in this paper indicate that the way forward in developing CNs is to bring together the experts from the areas of communication networks and AI. Communication networks are faced with great complexity challenges and several AI techniques proved to handle complexity well. Furthermore, AI is searching for areas of applications, and communication networks are underexploited in this respect. However, due to the vastness in Artificial Intelligence field, we hope to upgrade more in terms of Cognitive Networks and other methods & tools of AI.

References:

- [1] Nils J. Nilsson, Artificial Intelligence *A New Synthesis*.
- [2] R.W. Thomas, L.A. DaSilva, A.B. MacKenzie, Cognitive networks, in: Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, MD, USA, November 8–11, 2006.
- [3] J. Mitola, Cognitive Radio – An Integrated Agent Architecture for Software Defined Radio, Ph.D. Dissertation, Royal Institute of Technology, Kista, Sweden, May 8, 2000
- [4] Q. Mahmoud, Cognitive Networks – Towards Self-Aware Networks, John Wiley and Sons, 2007, ISBN 9780700999.
- [5] D.D. Clark, C. Partridge, J.C. Ramming, J.T. Wroclawski, A knowledge plane for the internet, in: Proceedings of the SIGCOMM 2004, Karlsruhe, Germany, August 26–29, 2004.
- [6] R.W. Thomas, Cognitive Networks, Ph.D. Dissertation, Virginia Polytechnic and State University, Blacksburg, VA, June 16, 2007.
- [7] J. Mitola, Cognitive Radio – An Integrated Agent Architecture for Software Defined Radio, Ph.D. Dissertation, Royal Institute of Technology, Kista, Sweden, May 8, 2000.
- [8] P. Balamuralidhar, R. Prasad, A context driven architecture for cognitive nodes, Wireless Personal Communications 16 (2008) 124– 110.
- [9] FCC, ET Docket No. 04-422, Notice of Proposed Rule Making and Order, December 2004. <<http://www.scribd.com/doc/112914/Federal-Communications-Commission-FCC04422A1>>.
- [10] QWL-QoS Ontology. <[http://www4.ntu.edu.sg/home6/PG0487868/ OWLQoSontology.html](http://www4.ntu.edu.sg/home6/PG0487868/OWLQoSontology.html)> (visited on August 2008).
- [11] P. Mahonen, M. Petrova, J. Riihijarvi, M. Wellens, Cognitive wireless networks: your network just became a teenager, in: Proceedings of the INFOCOM 2006, Barcelona, Spain, April 24–29, 2006.
- [12] B. Shepard, C. Matuszek, C.B. Fraser, W. Wechtenhiser, D. Crabbe, Z. Gundordu, J. Jantos, T. Hughes, L. Lefkowitz, M. Witbrock, D. Lenat, E. Larson, A knowledge-based approach to network security: applying Cyc in the domain of network risk assessment, in: Proceedings of the Innovative Applications of Artificial Intelligence Conference, Pittsburgh, PA, USA, July 9–14, 2006.
- [13] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, second ed., Prentice Hall, 2002, ISBN 0147901262.
- [14] J. Bradshaw, Software Agents, AAAI Press/The MIT Press, 1997, ISBN 0262622412.
- [15] P. Jackson, Introduction to Expert Systems, Addison-Wesley International Computer Science Series, 1986, ISBN 0201142246.

AUTHORS PROFILE

I am currently employed in Jazan University, Jazan, K.S.A Department of Computer Networks. My area of Research is Artificial Intelligence, Data Mining, Network Security. I have published few papers in International Journals. I wish to do research more into these fields.

A Performance Study on AES Algorithms

B.D.C.N.Prasad¹

Dept. of Computer Applications,
P V P Siddardha Institute of Technology
Vijayawada, India

P E S N Krishna Prasad²

Dept. of Computer Science & Engineering
Aditya Engineering College
Kakinada, India

P Sita Rama Murthy³

Dept. of Computer Science & Engineering
Sri Sai Aditya Institute of Science & Technology
Kakinada, India

K Madhavi⁴

Dept. of CSE
Dadi Institute of Technology
Anakapalli, India

Abstract— The Aim of this project is to find the performance comparative analysis of AES algorithms such as MARS, RC6, Rijndael, Serpent, Twofish algorithms in terms of speed, memory, time, encryption and decryption, key setup time, number of rounds, key sizes and also hardware considerations. Most of the AES algorithms, especially symmetric block ciphers, are based on the principle of substitution and transposition to encrypt a plain-text message and to produce a cipher-message. Those transformations are based on well-understood Mathematical problems using non-linear functions and linear modular algebra.

Implementation of cryptographic algorithms mainly uses bit-level operations and table look-ups. Bit-wise operators (XORs, AND/OR, etc.), substitutions, logical shifts and permutations are quite common operations. Such operations are well suited for their fast execution in hardware platforms. Furthermore, currently abundant memory resources in hardware platforms enhance encryption speed for the operations like substitution. These operators play an important role in analysis and comparison of the performance of the above mentioned AES algorithms, to evaluate simple, effective and efficient outcomes and also the information might be more secure.

Keywords-AES algorithms; Mars; RC6; Rijndael; Serpent; Twofish;

I. INTRODUCTION

Security is a broad topic and covers a multitude of sins, in its simplest form. It is concerned with making sure that nosy people cannot read, or worse yet, modify message intended for other recipients. It is concerned with people trying to access remote services that they are not authorized to use. Security also deals with people trying to deny that they sent certain message.

Network security problems can be divided roughly into four intertwined areas:

- Confidentiality,
- Authentication and Integrity control
- Denial of service

Cryptography, over the ages, has been practiced by many who have devised ad-hoc techniques to meet some of the information security requirements. The last twenty years have been period of transition as the discipline to a broader area. There are now several international scientific conferences denoted exclusively to cryptography and also and International Association for Crypto-logic Research (IACR), aimed at fostering research in the area.

There are two general types of cryptographic algorithms.

1. Symmetric algorithms.
2. Asymmetric algorithms.

The current Digital Encryption Standard (DES) does no longer satisfy the need for data security because of its short 56-bit key. Such short keys can today be broken by brute force attacks. We are looking for newer and more flexible algorithms to replace DES. Some of the candidates for the Advanced Encryption Standard (AES) are MARS encryption algorithm, RC6, Serpent, Rijndael, and Twofish. These are symmetric key block ciphers use 128 bit blocks and supports variable key sizes (from 128 to 1248 bits). These use addition and subtractions, S-boxes, fixed and data dependent rotations, and multiplications.

The final AES selection was made on the basis of several additional characteristics:

- computational efficiency and memory requirements on a variety of software and hardware, including smart cards
- flexibility, simplicity and ease of implementation

The existing system consisted of files with literally no file security standards like AES algorithms such as MARS, RC6, Rijndael, Serpent, and Twofish. AES algorithms are symmetric cipher algorithms which are far better than DES algorithms, since DES algorithms are limited key size with fixed number of blocks. So, we have chosen for finding the comparison of AES algorithms to provide the security for Data as well as networks and files. AES algorithms are to be

implemented due to the following factors against which several security measures had to be taken up:

1. Reading data
2. Manipulating and modifying data
3. Illegal use of files
4. Corrosion of data files
5. Distortion of data transmission

The main issue of (1) is *secrecy and confidentiality*. *Confidentiality* has always played an important role in diplomatic and military matters. Often Information must store or transferred from one place to another without being exposed to an opponent or enemy. Key management is also related to *Confidentiality*. This deals with generating, distributing and storing keys. Items (2-4) are primarily concerned with *reliability*. Often the expression *integrity* is used as a measure of genuineness of data. Also computer files and networks must be protected against intruders and Unauthorized. Item 5 is different aspect of the security of the information.

A. AES Algorithms

AES algorithms are symmetric cipher algorithms with variable key sizes and blocks, also with number of rounds to encrypt and decrypt the data than DES algorithms. There are numerous algorithms in AES. From them we have chosen the following algorithms for finding the performance analysis on time, memory, key sizes, key setup time, encryption, and decryption and so on.

The Chosen algorithms are as:

- MARS encryption algorithm
- RC6 Algorithm
- Rijndael Algorithm
- Serpent Algorithm
- Twofish Algorithm

TABLE I. GENERAL STRUCTURE

Cipher	Type	Rounds	Using
MARS	Extended Feistel	32	Variable Rotation, Multiplication Non Cryptic Rounds
RC6	Feistel	20	Variable Rotation, Multiplication
Rijndael	Square	10,12,14	
Serpent	SP Network	32	Bitslice
Twofish	Feistel	16	

1) Mode

No operational modes are currently defined for the AES cipher. The Cipher Block Chaining (CBC) mode is well-defined and well-understood for symmetric ciphers, and is currently required for all other ESP ciphers. This article specifies the use of the AES cipher and the other finalists in

CBC mode within ESP. This mode requires an *Initialization Vector* (IV) that is the same size as the block size. Use of a randomly generated IV prevents generation of identical cipher text from packets which have identical data that spans the first block of the cipher algorithm's block size.

The IV is XOR'd with the first plaintext block before it is encrypted. Then for successive blocks, the previous cipher text block is XOR'd with the current plaintext, before it is encrypted. For the use of CBC mode in ESP with 64-bit ciphers.

2) Key Size

Some cipher algorithms allow for variable sized keys, while others only allow specific, pre-defined key sizes. The length of the key typically correlates with the strength of the algorithm; thus larger keys are usually harder to break than shorter ones. This article stipulates that all key sizes MUST be a multiple of 8 bits.

The default key size that implementations MUST support 128 bits. In addition, all of the ciphers accept key sizes of 192 and 256 bits.

TABLE II. KEY SIZES

Algorithm	Key Sizes(bits)	Default
MARS	128 – 448*	128
RC6	Variable up to 2040	128
Rijndael	128,192,256	128
Serpent	Variable up to 256**	128
Two fish	Variable up to 256***	128

MARS key lengths must be multiples of 32 bits.

** Serpent keys are always padded to 256 bits. The padding consists of a "1" bit followed by "0" bits.

*** Twofish keys, other than the default sizes, are always padded with "0" bits up to the next default size.

3) Weak Keys

Some cipher algorithms have weak keys or keys that MUST not be used due to their interaction with some aspect of the cipher's definition. If weak keys are discovered for the AES or any of the other finalists, then weak keys SHOULD be checked for and discarded when using manual key management. When using dynamic key management, weak key checks SHOULD NOT be performed as they are seen as an unnecessary added code complexity that could weaken the intended security.

4) Block Size and Padding

All of the algorithms described in this document use a block size of sixteen octets (128 bits), mandatory for the AES. Some of the algorithms can handle larger block sizes as well. Padding is required by the algorithms to maintain a 16-octet (128-bit) blocksize. Padding MUST be added, such that the data to be encrypted has a length that is a multiple of 16 octets. Because of the algorithm specific padding requirement, no additional padding is required to ensure that the cipher text terminates on a 4-octet boundary (i.e. maintaining a 16-octet blocksize guarantees that the ESP Pad

Length and Next Header fields will be right aligned within a 4-octet word

5) Rounds

This variable determines how many times a block is encrypted. While this variable MAY be negotiated, a default value MUST always exist when it is not negotiated.

Algorithm	Negotiable	Default of Rounds
MARS	yes	32
RC6	yes	20
Rijndael	yes	10,12,14
Serpent	yes	32
Twofish	yes	16

B. MARS Algorithm

MARS is a shared-key block cipher that works with a block size of 128 bit and a variable key size. The algorithm is a type-3 Feistel network which is word (32 bit) oriented. The word orientation should bring a performance for software implementations on most computer architectures available today. A fully optimized implementation is expected to run at 100Mbit/second and hardware can achieve an additional 10x speedup factor.

1) Operations

MARS algorithm uses a big variety of different operations:

Additions, subtractions and xors: These simple operations are used to mix data and key values together. Because xors are interleaved with additions and subtractions these operations do not commute with each other.

Table look-up: Similar to the S-boxes in DES has also MARS cipher a table look-up. It uses a single table of 512 32-bit words, simple called S-box. A problem of the table look-up is the slow software implementation (at least 3 instructions per look-up). That's why S-box look-up is only used sparsely in MARS where fast avalanche of the key bits is needed.

Fixed rotations: Data-dependent rotations: Data dependent rotations may lead to differential weaknesses. This problem is solved in MARS by combining these rotations with multiplications.

Multiplications: All multiplications in MARS are modulo 232 which suits most modern computer architectures. Multiplications used to be a problem in cryptographic algorithms because they were slow. Today is this no longer the case. Most architecture can complete a multiplication in 2 clock cycles. MARS algorithms uses 16 multiplications per block. This should not be a big deal. For hardware realizations we have the problem that a multiplier needs much more chip-space than adders or logical units.

C. Comparison with other AES Candidates

There are 4 other candidates for AES in the last round. So they are all 128 bit block ciphers with variable key length from 128 bit to at least 192 bit. All designs claim to be secure against all known attacks like differential, linear, known plaintext or cipher text and other attacks.

1) RC6

RC6 is the submission of MIT (Massachusetts Institute of Technology) and the RSA-Laboratories. Similar to MARS it splits the 128 bit blocks into four words in the algorithm, but the algorithm is designed in a way that you can easily change to two 64 bit words in newer architectures. RC6 is also a Feistel network. It uses the same type of operations except from look-up tables and fixed rotations. The algorithm is more flexible than MARS about blocksize and number of rounds. The AES submission is optimized for 128 bit blocks and 20 rounds. Several performance test showed that RC6 is slower than MARS for encryption and for the key setup. But it uses less memory because there are no look-up tables.

2) Rijndael

Rijndael is the submission of the Belgium Proton World Int. and the Katholieke Universities Leuven, Belgium. This algorithm is quite different from MARS. It works with Galois Field GF(128) arithmetic and handles the input blocks as matrices of bytes. They define several operations to these matrices as ByteSub, ShiftRow, MixColumn and AddRoundKey. For detailed information about these operations consult [Rijndael99]. Several combinations of these operations define a round. Depending on the key length which is in the range from 128 to 256 bits a fixed number of rounds has to be executed. This cipher is not a Feistel network. Several performance tests showed that Rijndael is about the same speed in encryption and decryption as MARS. But the key expansion for keys of the same length is significant slower.

3) Serpent

Serpent is a submission from three universities (Cambridge University, England; Technion, Haifa, Israel; University of Bergen, Norway). Therefore it's the only algorithm where no company stands behind. The algorithm is pretty similar to DES, it uses permutations, xors, fixed rotations and shifts and constant table look-up's. The first version of the algorithm even used the same S-boxes as DES. The key can vary from 128 to 256 bit. The algorithm works internally also with 4 words as RC6 and MARS. Performance tests that the encryption of Serpent is about 25% faster than the MARS encryption. But the key expansion is significant slower. An implementation of Serpent also uses a lot of memory because of the look-up tables.

4) Twofish

Twofish is the submission from a company called Counterpane. It is a 16 round Feistel cipher that works with key dependent 8x8 bit look-up tables, 4 by 4 matrices over the Galois field GF(128), a pseudo-Hadamard transform, permutations and rotations. The detailed description of these functions can be found in [Twofish]. The key length varies also from 128 bit to 256 bit as in most other AES candidates. Performance tests showed that the encryption speed of Twofish is about the same as for MARS, but the Twofish key setup is significant faster.

D. Performance Analysis

The performance analysis can be done with various measures such as speed comparison with encryption and decryption cycles, key setup and key initialization, analysis of various key sizes and fair speed/security comparisons. The performance analysis will be presented in the form of tables and figures below.

1) Speed Comparisons

TABLE III. SPEED

Cipher	Encrypt (Cycles)	Decrypt (Cycles)	Key Setup		Init
			Encrypt	Decrypt	
MARS	1600	1580	4780	5548	18
RC6	1436	1406	5186	5148	30
Rijndael	1276	1276	17742	18886	28
Serpent	1800	2102	13154	12648	14
TwoFish	1254	1162	18846	18634	20

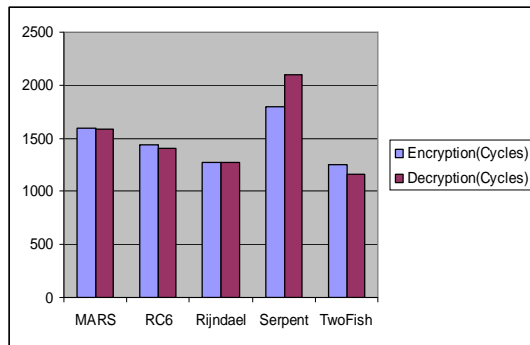


Figure 1. Graph for Encryption and Decryption (Cycles)

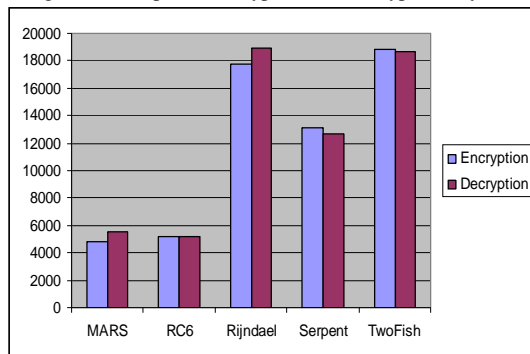


Figure 2. Graph for Key setup Encryption and Decryption

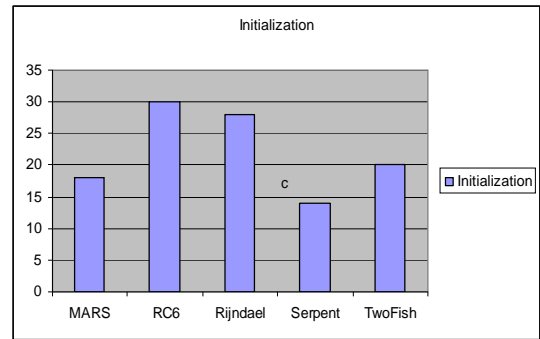


Figure 3. Key Initialization

2) Analysis on various Key Sizes

a) Encryption

TABLE IV. ENCRYPTION

Algorithm	Encry128	Encry192	Encry256
MARS	3738	3707	3733
RC6	4698	4740	4733
Rijndael	4855	4664	4481
Serpent	1843	1855	1861
Twofish	1749	1749	1744

b) Decryption

TABLE V. DECRYPTION

Algorithm	Encry128	ency192	ency256
MARS	3965	3965	3936
RC6	4733	4698	4740
Rijndael	4819	4624	4444
Serpent	1873	1897	1896
Twofish	1781	1775	1761

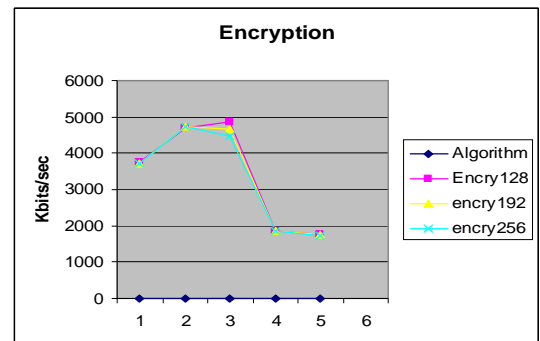


Figure 4. Encryption

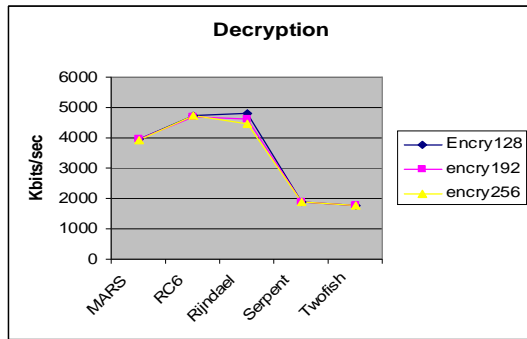


Figure 5. Decryption

c) Fair Speed/ Security Comparisons

TABLE VI. FAIR SPEED/ SECURITY COMPARISONS

Cipher	Original (cycles)	Rounds	Minimal Rounds	Time (Cycles)
MARS	1600	32	20	1000
RC6	1436	20	20	1436
Rijndael	1276	10	8	1021
Serpent	1800	32	17	956
Twofish	1254	16	12	940

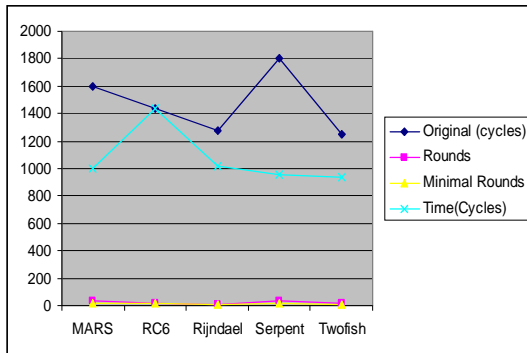


Figure 6. Fair speed / security comparisons

E. Conclusion

A performance comparison can be made among various AES Algorithms such as MARS, RC6, Rijndael, Serpent, Twofish. The Performance analysis reports were presented in the specified contents. It is concluded that all the above specified algorithms have almost similar speed rate and timings while using java tool for execution of these

algorithms. But MARS is one among the chosen algorithms is some what better as considered reports.

It won't be an easy decision to choose one of the finalists as AES. There is no known weakness in all these algorithms, so other factors as performance, needed hardware or flexibility must be used for the decision. MARS cipher is for sure a good candidate. It has the largest available key length of all of them and it is expandable to larger block sizes than 128 bit. Another advantage of MARS is that it comes from a well known company that is in this business for a long time which means they have a lot of experience and have proven their trustworthiness.

F. References

- [1] Cryptography and Network Security -"William Stallings" ,Third Edition.
- [2] The Laws of Cryptography with JAVA Code -"Neal R.Wagner".
- [3] **MARS**: C.Burwick, D.Coppersmith, E.D'Avignon, R.Gennaro, S.Halevi, C.Jutla, S.Matyas, L.O'Connor, M.Peyravian, D.Safford, N.Zunic, "MARS - a candidate cipher for AES", IBM Corporation, September 1999.
- [4] TweakIBM99 - Shai Halevi, "Detailed discussion of the MARS "tweak" for Round 2", IBM Corporation, Mai 1999.
- [5] **RC6**: Ronald L. Rivest, M.J.B. Robshaw, R. Sidney, Y.L. Yin, "The RC6 Block Cipher", M.I.T. Laboratory for Computer Science, RSA Laboratories.
- [6] **Rijndael**: Joan Daemen, Vincent Rijmen, "AES Proposal: Rijndael", Proton World Int.l, Belgium, Katholieke Universiteit Leuven, Belgium, September 1999.
- [7] **Serpent**: Ross Anderson, Eli Biham, Lars Knudsen, "Serpent: A Proposal for the Advanced Encryption Standard", Cambridge University, England; Technion, Haifa, Israel; University of Bergen, Norway.
- [8] **Twofish**: Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, Niels Ferguson, "Two sh: A 128-Bit Block Cipher", Counterpane Systems, University of California Berkeley.
- [9] E Biham, "A Note Comparing AES Candidates, NIST,1999.
- [10] P. Preneel, V Rijmen and A Bosselaers, " Principles and Performance of Cryptographic Algorithms", Dr. Dobb's journal.
- [11] B. Schneier, J Kelsey, D. Whiting, D Wagner, C. Hall and N Ferguson, "Performance Comparison of the AES Candidate conference,1999.

AUTHORS PROFILE



Dr. B D C N Prasad, currently is a Professor & Head of Department of Computer Applications at Prasad V. Potluri Siddardha Institute of Science and Technology, Vijayawada, Andhra Pradesh, India. He received Ph.D. in Applied Mathematics from Andhra University, Visakhapatnam, India in 1984. His research interests

includes Machine Intelligence, Data Mining, Rough Sets and Information Security in Computer Science and Boundary value problems and Fluid Dynamics in Mathematics. He has several publications in mathematics and computer science in reputed national and international journals. He is a member of ISTAM , ISTE and also he is a national executive member of Indian Society for Rough Sets.



Mr. P E S N Krishna Prasad, currently is a Research Scholor under the guidance of Dr. BDCN Prasad in the area of Machine Intelligence and Neural Networks. He is working as Associate Professor in the Department of CSE, Aditya Engineering College, Kakinada, Andhra pradesh, India. He is a member of ISTE. He has presented and published papers in several national and International conferences and journals. His areas of interest are Artificial Intelligence, Neural Networks and Machine Intelligence.



Mr. P Sita Rama Murty, currently is a Research Scholor, in the area of ATM networks and Information Secuirty. He is working as Assistant Professor in the department of CSE, Sri Sai Aditya Institute of Science and Technology, Kakinada, Andhra Pradesh, India .

Hybrid Fingerprint Image compression and Decompression Technique

*Dr.R.Seshadri, ,B.Tech.,M.E,Ph.D

**Yaswanth Kumar.Avulapti ,M.C.A, M.Tech, (PhD)

***Dr.M.Usha Rani M.C.A,PhD

*Director, S.V.U.Computer Center S.V.University, Tirupati

**Research Scholar, Dept of Computer Science, S.V.University, Tirupati

***Associate Professor, Dept. of Computer Science,, SPMVV, Tirupati

Abstract

In this paper a biometric authentication system based on Fingerprint. A fingerprint is the representation of the epidermis of a finger. It consists of a pattern of interleaved ridges and valleys.

Like every thing in the body fingerprint ridges form through a combination of genetic and environmental factors. Finger prints are fully formed at about seven months of fetus development. Fingerprint ridges don't change throughout the life of an individual except incase of accidents such as cuts on the fingertip (or) burns on the fingertip. In this paper we proposed a hybrid model to compress the fingerprints.

Keywords:Biometrics,Enrollment Authentication, compression, Decompression.

Introduction

The term Biometrics is derived from the Greek word bio (life) and metrics (to measure). Basically it is a method of identifying a person based on his/her

physiological or behavioral characteristics such as Fingerprints, Iris, Face, Hand geometry, Retinal scan,,DNA,Signature,Key Stroke, Voice,Gait,Ear,Palm print, Dental radiographs. Among all the biometric techniques, fingerprint recognition is the most popular method and is successfully used in many applications

Biometrics is a rapidly evolving wonderful technology which has been widely used in forensics applications such as criminal identification and prison security. Biometrics(Fingerprints) can be used to prevent unauthorized access to a computer.

The main objective of the fingerprint image compression is to reduce the number of bits as much as possible by keeping the resolution and quality of the fingerprint while decompressed as same as to the original fingerprint image.

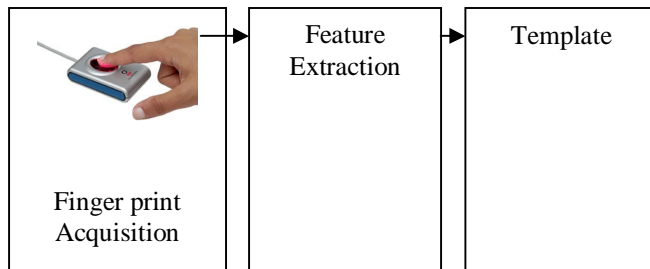
Fingerprint Identification System

A finger print system works in two different modes they are Enrollment mode and Authentication mode as shown in figure.1

Enrollment mode in which fingerprint system is used to identify and collect the related information about the person and his/her fingerprint image.

Authentication mode in which fingerprint system is used to identify the person who is declared to be him/her.

Enrollment Mode



Authentication Mode

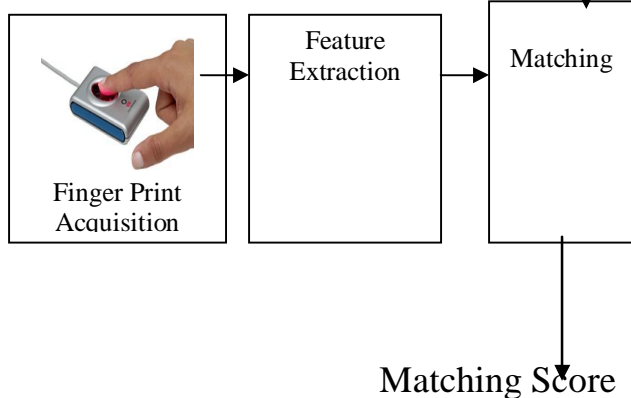


Fig.1.Enrollment and Authentication of a Fingerprint system

Need for fingerprint compression

Let us see this scenario the spatial resolution of the cinefilm is generally assumed to be equivalent to a digital matrix of at least 1000 by 1000 pixels, each with up to 256 gray levels (8 bit or one byte) of contrast information.

The Figure.2 derives from this principal parameter some examples for requirements on digital image communication and archival in a catheterization laboratory with low to medium volume.

	Film	Digital
Spatial resolution	4 linepairs/mm	1024 by 1024 pixels
Data capacity per image		1MB
Data rate	30 images per seconds	30 MB/Sec
Media	one film	4-CDs

Fig.2.Replacement of cine film by digital Imaging with high resolution

In this scenario the enormous data rate of 30 Megabyte per second has to be supported. This is much faster than even advanced ATM networks (offering less than 20 MB/s or 160 Mbit/s).

Looking for existing off-line media real-time display from CD-R would require a CD-R player with a data rate of 200X, while the fastest players available presently deliver 50X (1X stands for a data rate of 150 KB/sec). The total amount of data required in this scenario is even more frightening. To store these images and make them available over network (e.g. the internet), compression techniques are needed.

Steps involved in Fingerprint compression

- 1) Specifying the rate (bits available) and distortion parameters for the target fingerprint image
- 2) Dividing the fingerprint data into various classes based on their importance
- 3) Dividing the available bit budget among these classes such that the distortion is minimum.
- 4) Quantize each class separately using the bit allocation
- 5) Encode each class separately and write to the file

Steps involved in Fingerprint decompression

- 1) Read the fingerprint quantized data from file using the decoder
- 2) Dequantize the fingerprint data (Reverse of step 4)
- 3) Rebuild the Fingerprint Image (Reverse of step 2)

Proposed Fingerprint Image Compression System

We proposed a Fingerprint image compression system in which the fingerprint images are compressed by Lossless compression technique called Run length Encoding and Decompressed by Huffman coding. It is also called as Hybrid technique.

A Fingerprint compression system consists of two blocks namely Encoder and Decoder as shown in the fig.3

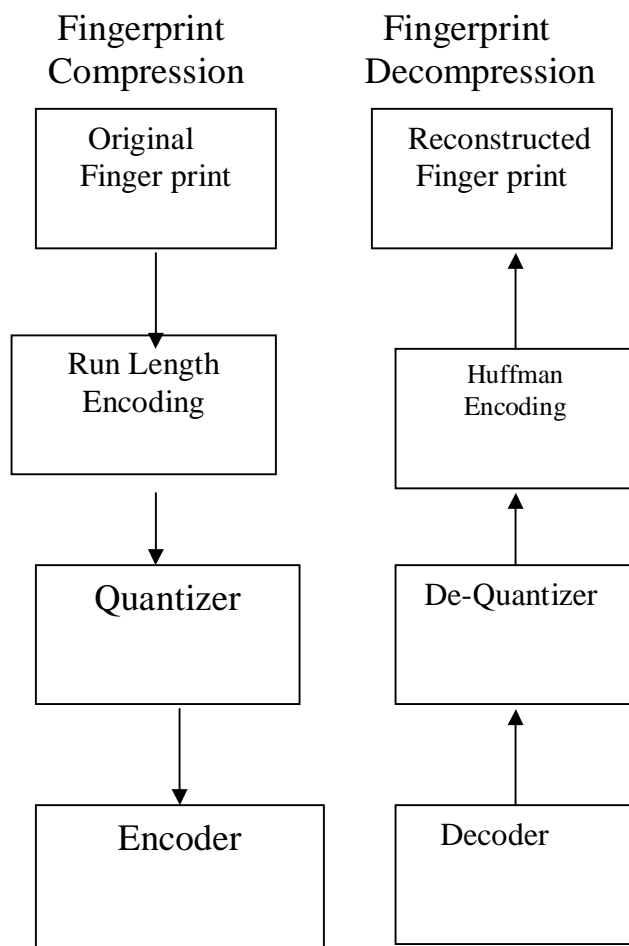


Fig.3. Fingerprint image compression using Run length Encoding & Decompressed using Huffman coding

Steps in Proposed Fingerprint compression

- 1) Specifying the bits available and distortion parameters for the target fingerprint image
- 2) Dividing the fingerprint data into various classes based on their importance
- 3) Dividing the available bit budget among these classes such that the distortion is minimum.
- 4) Quantize each class separately using the bit allocation
- 5) Encode each class separately using an Run Length encoder and write to the file

Steps involved in Fingerprint decompression

- 1) Read the fingerprint quantized data from file using the Huffman decoder
- 2) Dequantize the fingerprint data (Reverse of step 4)
- 3) Reconstructed the Fingerprint image (Reverse of step 2)

In the Run length encoding the Fingerprint images with repeating grey values along rows or columns can be compressed by storing "runs" of identical grey values as shown in the fig.4 & fig 4.b

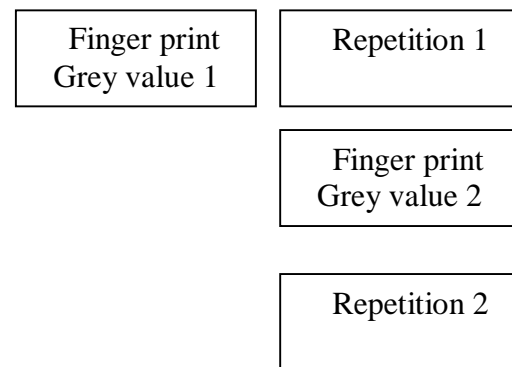


Fig.4.a Run length encoding the Fingerprint Images

This is a very simple fingerprint compression method used for chronological data. The run length encoding replaces the sequences of identical pixels called the runs

by tiny symbols. The run length code for fingerprint image is represented by a sequence of {A,B} where a is the intensity of the pixel and B is the number of consecutive pixels with the intensity A as shown in the figure .If both A and B are represented by 8-bits this span of 12 pixels is coded using eight bytes yielding a compression ratio 1:5.

For Example

23,23,23,23,23,67,67,89,89,78,78

{23,5}{67,2}{89,2}{78, 2}

Fig.4.bRun length encoding the Fingerprint Images

Huffman encoding is performed that is mapping of the code words to the corresponding symbols will result in a compressed data. The original image is reconstructed i.e. decompression is done by using Huffman decoding.

Generate a tree equivalent to the encoding tree. Read input character wise and left table until last element is reached in the table. Output the character encodes in the leaf and return to the root, and continues until all the codes of corresponding symbols.

Advantages of Fingerprint Compression:

- a) It reduces the storage space at the time of processing the fingerprints
- b) It not only reduces the storage requirements but also overall execution time
- c) It also reduces the probability of transmission errors since fewer bits are transferred
- d) It also provides a level of security against criminal monitoring

Conclusion

This paper presents different steps involved in the development of fingerprint authentication system. The proposed fingerprint image compression and decompression technique uses both the Run length encoding and Huffman coding. This hybrid fingerprint compression and decompression technique are good for certain applications like the security technologies. These two compression and decompression techniques are lossless ones. Using this hybrid technique which leads to less storage of memory and reducing the execution time.

References

- 1.A. K. Jain,Patrick Flynn,Arun A.Ross . “Handbook of Biometrics”.
2. The Henry Clas sification System Copyright © 2003 International Biometric Group
- 3.Compression Using Fractional Fourier Transform A Thesis Submitted in the partial fulfillment of requirement for the award of the degree of M.E Electronics &Communication. By Parvinder Kaur.
4. RL-Huffman Encoding for Test Compression and Power Reduction in Scan Applications-MEHRDAD NOURANI and MOHAMMAD H. TEHRANIPOUR, The University of Texas at Dallas
- 5.A.B.Watson,“Image Compression using the DCT” ,Mathematica Journal, 1995,pp.81-88.
6. DAVID A. HUFFMAN, Sept. 1991, profile Background story: Scientific American, pp. 54-58.
7. Efficient Huffman decoding by MANOJ Aggarwal and Ajai Narayana A NEW LOSSLESS METHOD OF IMAGE COMPRESSION AND DECOMPRESSION USING HUFFMAN CODING TECHNIQUES by JAGADISH H. PUJAR, 2LOHIT M. KADLASKAR Faculty, Department of EEE, B V B College of Engg. & Tech. India
- 2 Student, Department of EEE, B V B College of Engg. & Tech. India
8. D. Monro and B. Sherlock. Space-frequency balance in biorthogonal wavelets. transactions

of the IEEE Int.Conf. on Image Processing, 1:624{627, 1997.

9. A. Said and W. Pearlman. A new fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Transactions on Circuits and Systems for Video Technology, 6:243{250, June 1996.

10. W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets.

Authors Profile



Dr.R.Seshadri was born in Andhra Pradesh, India, in 1959. He received his **B.Tech** degree from Nagarjuna University in 1981. He received his **M.E** degree in Control System Engineering from PSG College of Technology, Coimbatore in 1984. He was awarded with **PhD** from Sri Venkateswara University, Tirupati in 1998. He is currently Director, Computer Center, S.V.University, Tirupati, India. He has Published number of papers in national and international conferences, seminars and journals. At present 12 members are doing research work under his guidance in different areas



Mr.YaswanthKumar .Avulapati received his **MCA** degree with **First class** from Sri Venkateswara University, Tirupati. He received his **M.Tech** Computer Science and Engineering degree with **Distinction** from

Acharya Nagarjuna University, Guntur.He is a research scholar in S.V.University Tirupati,³⁸ Andhra Pradesh.He has presented number of

papers in national and international conferences, seminars. He attends Number of work shops in different fields.



Dr. M. Usha Rani is an Associate Professor in the Department of Computer Science and HOD for CSE&IT, Sri Padmavathi Mahila Viswavidyalayam(SPMVV Womens' University),

Tirupati. She did her Ph.D. in Computer Science in the area of Artificial Intelligence and Expert Systems. She is in teaching since 1992. She presented more than 34 papers at National and Internal Conferences and published 19 articles in national & international journals. She also written 4 books like Data Mining - Applications: Opportunities and Challenges, Superficial Overview of Data Mining Tools, Data Warehousing & Data Mining and Intelligent Systems & Communications. She is guiding M.Phil and Ph.D. in the areas like Artificial Intelligence, DataWarehousing and Data Mining, Computer Networks and Network Security etc.

Punctured Self-Concatenated Trellis Codes with Iterative Decoding

Labib Francis Gergis

*Misr Academy for Engineering and Technology
Mansoura City, Egypt*

Abstract-A special concatenated code structure called self-concatenated trellis code (SCTC) is presented. This scheme based on only one recursive convolutional code(RSC), followed by a mapping modulator. The union bounds of SCTC are derived for communications over Additive White Gaussian Noise (AWGN) and Rayleigh fading channels. Asymptotic results for large interleavers are extended to M-ary bandwidth efficient modulation schemes by puncturing process. The combination of self-concatenated codes with powerful bandwidth-efficient component codes leads to a straightforward encoder structure, and allows iterative decoding. The scheme has been investigated for 4-PSK, 8-PSK, 16-PSK, and 16-QAM modulation schemes with varying overall bandwidth efficiencies. The choice based on the rate of RSC and puncturer encoder component.

key words ;Self-Concatenated codes, trellis-coded modulation, uniform interleaved coding, convolutional coding, iterative decoding

1. INTRODUCTION

Trellis coded modulation (TCM) [1] was originally proposed for transmission over AWGN and fading channels due to its attractive bandwidth efficiency.

Concatenated trellis-coded modulation is an alternative to TCM. Different approaches to concatenated trellis-coded modulations were presented in [2], and [3]. The main principle in the concatenated coding schemes is to use two codes in series (or parallel) joined through one or more interleavers. This means that the information sequence is encoded twice, the

second time after a scrambling of the information bits.

Concatenated trellis codes are classified as serially concatenated convolutional codes (SCCC), these codes were analyzed in [4]. Using the same ingredients, another type of concatenated codes named parallel concatenated convolutional codes (PCCC), was described in [5]. A third choice is defined as a hybrid concatenation of convolutional codes (HCCC) was described in [4] and [6]. Self-concatenated convolutional codes proposed in [7], [8], and [9] constitute another attractive iterative detection aided code-family for their low complexity, since they invoke only a single encoder and a single decoder.

Puncturing is the process of deleting some parity bits from the codeword according to a puncturer code rate. The redundant bits in coding decrease the bandwidth efficiency. Puncturing increases code rate without increasing complexity and decreases free distances of code. The advantage of punctured codes for binary transmission is that the encoders and decoders for the entire class of codes constructed easily by modifying the single encoder and decoder for the rate $1/2$ binary convolutional code from which the high rate punctured code was derived [10].

The construction of self-concatenated trellis codes (SCTC) is described in section 2. Section 3, derives analytical upper bounds to the bit-error probability of SCTC using the concept of uniform interleavers. Factors that affect the performance of SCTC are described in section 4. Finally results for some examples depicted in section 4, have been stated in section 5.

2. SCTC MODEL

The basic concept of self-concatenated scheme is shown in Figure. 1, the input bit sequence $\{b_1\}$ of the self-concatenated encoder is interleaved to yield the bit sequence $\{b_2\}$. After the parallel-to-serial (P/S) conversion, the information sequence is defined as $b^{(1)} = \{b_{1,1} b_{2,1} b_{1,2} b_{2,2} \dots\}$. The resultant bit sequences are input to a recursive systematic convolutional (RSC) encoder. At the output of the encoder the interleaved bit sequence is punctured. The encoder output is composed of the combined systematic bit sequence and parity bit sequence.

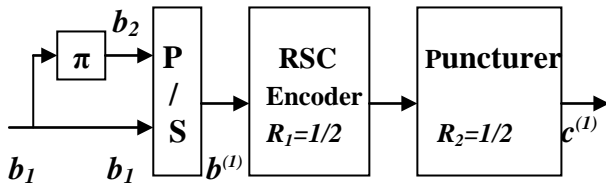


Fig 1. The Self-Concatenated Code Encoder

The overall code rate, R , can be derived based on [9] as:

$$R = R_1 / 2 R_2 = (1/2) / 2 (1/2) = 1/2 \quad (1)$$

It can be observed that different codes can be designed by changing R_2 .

3. PERFORMANCE OF SELF-CONCATENATED TRELLIS CODES

Consider a linear block code C with code rate R , and minimum distance h_m . An upper bound on the conditional bit-error probability of the block code C over AWGN channels, assuming coherent detection, maximum likelihood decoding, can be obtained in the form [4]

$$P_b(e/\rho) \leq \sum_{h=d_{min}}^{N/R} \sum_{w=1}^N (w/N) A_{w,h}^c \cdot Q \sqrt{R h (E_b / N_o)} \quad (2)$$

where E_b/N_o is the bit energy to noise density ratio, $A_{w,h}^c$ for block code C represents the number of codewords of the block code with output weight h associated with an input sequence of weight w , and N is the size of the interleaver. The $A_{w,h}^c$ is the input-output weight coefficient (IOWC). The function $Q(\sqrt{2R h E_b/N_o})$ represents the pairwise error probability which is a monotonic decreasing function of the signal to noise ratio and the output weight h .

For a fading channels, assuming coherent detection, and perfect Channel State Information (CSI), the conditional pairwise error probability is given by

$$Q \left(\sqrt{2R E_b/N_o \sum_{i=1}^h \rho_i^2} \right) \quad (3)$$

The fading samples ρ are independent identically distributed (i.i.d.) random variables with Rayleigh density of the form

$$f(\rho) = 2 \rho e^{-\rho^2} \quad (4)$$

The structure of a SCTC, as shown in Figure .1, is composed of $q-1$ interleavers each of size N bits, and a single systematic recursive trellis code C with rate $(bq/bq+1)$, where only the $b+1$ outputs of the encoder are mapped to 2^{b+1} modulation levels.

The average input-output weight coefficients $A_{w,h}^c$ for SCTC with $q-1$ interleavers can be obtained by averaging equation (2) over all possible interleavers. A uniform interleaver is defined as a probabilistic device that maps a given input word of weight w into all its distinct $\begin{bmatrix} N \\ w \end{bmatrix}$ permutations with equal probability $1/\begin{bmatrix} N \\ w \end{bmatrix}$.

Thus, the expression for IOWC of SCTC is derived as [7]

$$A_{w,h}^c = \frac{A_{w,w,\dots,w,h}^c}{\begin{bmatrix} N \\ w \end{bmatrix}} \quad (5)$$

where $A_{w,w,\dots,w,h}^c$ is the number of code words of the trellis encoder of weight h , which is determined in [5], and

$$\binom{N}{w} \approx \frac{N^w}{w!} \quad (6)$$

Substituting equation (6) in equation (2) yields [3]

$$P_b(e/\rho) = B_m N^{-q+1} Q \left(\sqrt{2 R h_m E_b / N_o} \right) \quad (7)$$

where the constant B_m is independent of N , and is derived in [8], and h_m is the minimum Euclidean distance of the SCTC scheme.

4. SCTC: PERFORMANCE FACTORS

It is shown from equation (7), that there are many factors that affect the performance of SCTC. The most influential parameter is the interleaver size N . The bit error probabilities for self-concatenated trellis code with overall rate $R=1/2$, is shown in Fig. 2, with various interleaver lengths $N=10, 50, 100, 200$, and 300 are plotted versus the signal-to-noise ratio E_b/N_o . The systematic and parity bits, b_o and b_1 , are mapped to 4-ary Phase Shift Keying (QPSK) modulation. The figure shows the beneficial gain that can be achieved through increasing N .

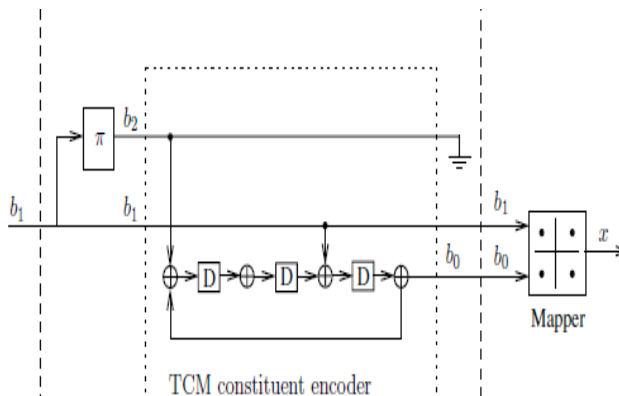


Fig. 2. Self-Concatenated Trellis Encoder with rate $R = 1/2$

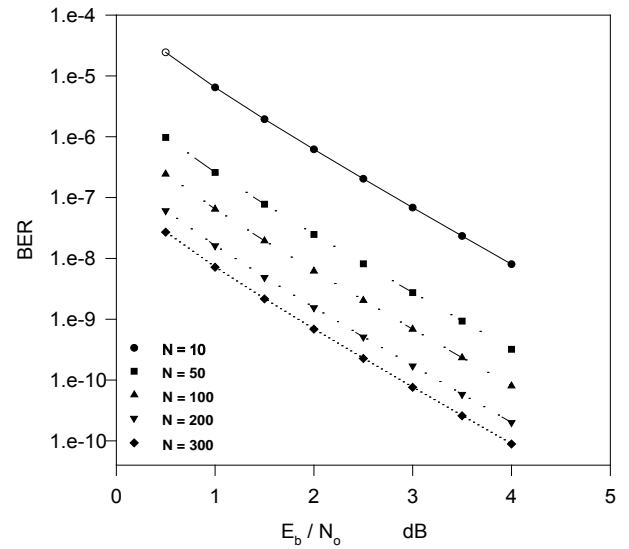


Fig. 3. Upper Bounds to the Bit Error Probability for SCTC with QPSK using different Interleaver Lengths

Applying the upper bound of equation (7), we obtain the results reported in Fig. 3. It is also clear from equation (7) that, the minimum Euclidean distance of the SCTC code (h_m) is another main parameter affecting the performance of SCTC. Different values of h_m could be obtained by a variety of modulation schemes. Puncturing is used in order to increase the achievable bandwidth efficiency. Different codes could be designed by changing the rates R_1 and R_2 . The output of the encoder is then mapped to the Gray-code mapping function. The various coding schemes considered in this paper are characterized in Table 1, that defines both R_1, R_2 , the overall code rate R , and the associated mapped modulation scheme to R . The BER versus E_b/N_o performance curves of the various QPSK, 8-PSK, 16-PSK, and 16-QAM are shown in Fig. 4.

R_1	R_2	R	Modulation Scheme
$1/2$	$1/2$	$1/2$	QPSK
$1/3$	$1/4$	$2/3$	8-PSK
$1/3$	$2/3$	$1/4$	16-PSK 16-QAM

Table 1.

Various Modulation Schemes Obtained from Varying R_1 and R_2 .

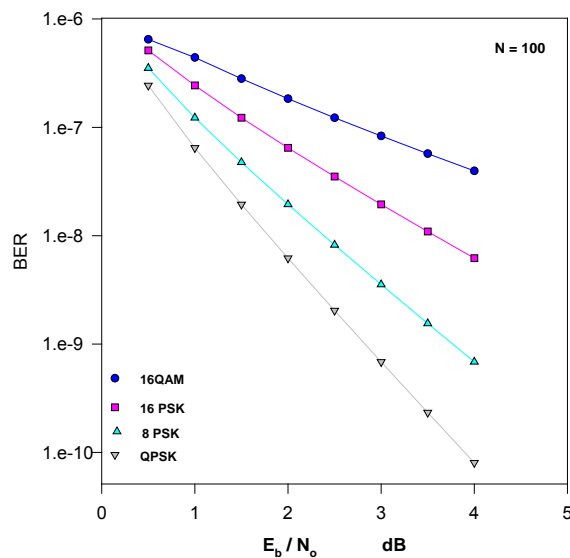


Fig. 4. Upper Bounds to the Bit Error Probability for SCTC versus Different Modulation Schemes

The choice of decoding algorithm and number of decoder iterations also influences performance.

A functional diagram of the iterative decoding algorithm for SCTC is presented in Fig. 5.

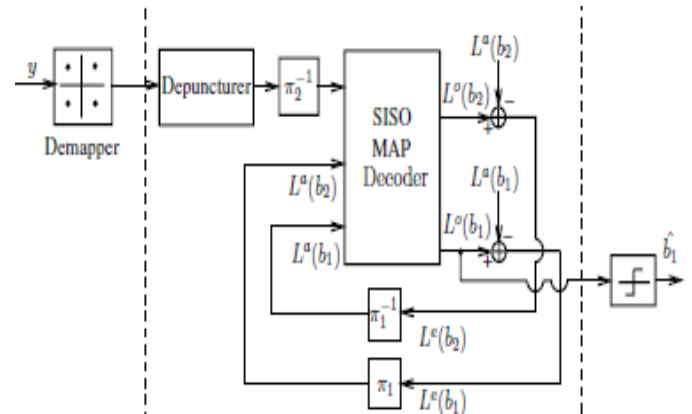


Fig 5. Self-Concatenated Trellis Decoder

The decoder is a self-concatenated scheme using a soft-input soft-output (SISO) *maximum a posteriori probability* (MAP) algorithm [9]. It first calculates the extrinsic loglikelihood Ratio (LLR) of the information bits, namely $L_e(b_1)$ and $L_e(b_2)$. Then they are appropriately interleaved to yield the a priori LLRs of the information bits, namely $L_a(b_1)$ and $L_a(b_2)$, as shown, in Fig. 5. Self-concatenated decoding proceeds, until a fixed number of iterations is reached.

The performance of SCTC with QPSK modulation schemes considered are shown in Fig. 6. The SCTC has an overall rate $R = 1/2$, the interleaver length N of this code = 100 bits. The performance after various numbers of iteration is shown. It is clear that performance improves as the number of decoder iterations increases.

5. CONCLUSIONS

In this paper, a channel coding scheme (SCTC) that is bandwidth efficient and allows iterative decoding of codes built around punctured codes together with higher order signaling.

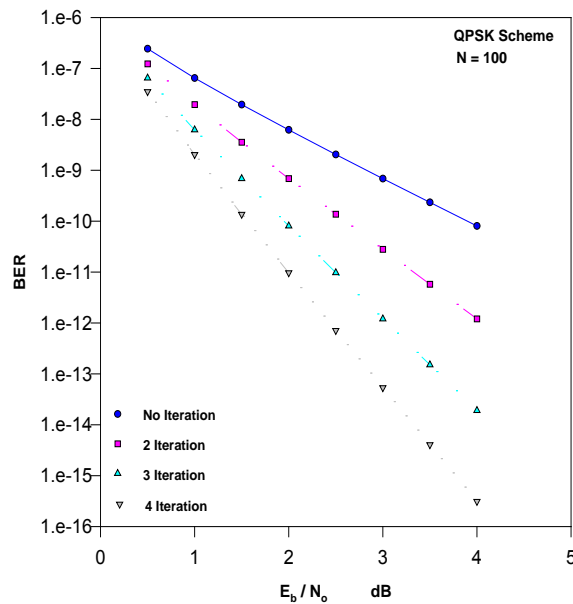


Fig. 6. Upper Bounds to the Bit Error Probability for SCTC versus Different Decoding Iterations

The SCTC schemes consists of binary RSC codes and different puncturing rates. The puncturer is used to increase the achievable bandwidth efficiency. A search for good rates was performed, taking into account the puncturing at the transmitter. It is also demonstrated the significant in the performance and the decrease of the bit error rate and probability of errors to SCTC within increasing: the interleaver size N , and the number of decoder iterations

REFERENCES

- [1] G. Ungerboeck, "Channel coding with multilevel phase signaling," *IEEE Trans. Inf. Th.*, Vol. 25, pp. 55-67, Jan. 1982
- [2] F. Brannstrom, A. Amat, and L. Rasmussen, "A General Structure for Rate-Compatible Concatenated Codes," *IEEE Trans on Communication Letters*, Vol. 11, pp. 437-439, May 2007.
- [3] A. Amat, G. Montorsi, and S. Benedetto, "New High-rate Convolutional Codes for Concatenated Schemes", *Proceeding of IEEE International Conference on Communication, ICC 2002*, Vol. 3, pp. 1661-1666, April 2002.
- [4] D. Divsalar, and F. Pollara, "Serial and Hybrid Concatenated Codes with Applications," *International Symposium on Turbo Codes and Related Topics*, Brest, France 1997
- [5] S. Benedetto, and G. Montorsi, "Unveiling Turbo Codes: Some Results On Parallel Concatenated Coding Schemes," *IEEE Transactions on Information Theory*, Vol. 42, No. 2, March 1996
- [6] A. Amat, and E. Rosnes, "Good Concatenated Code Ensembles for the Binary Erasure Channel", *IEEE Journal on Selected Areas in Communications*, Vol. 27, No. 6, pp. 928-943, August 2009.
- [7] S. Ng, M. Butt, and L. Hanzo, "On the Union Bounds of Self-Concatenated Convolutional Codes", *IEEE Signal Processing Letters*, Vol. 16, No. 16, No. 9, September 2009.
- [8] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Self-Concatenated Codes with Self-Iterative Decoding for Power and Bandwidth Efficiency" *International Symposium on Information Theory, ISIT 1998*, Cambridge, MA, USA, August 1998.
- [9] M. Butt, R. Riaz, S. Ng, and L. Hanzo, "Distributed Self-Concatenated Codes for Low-Complexity Power-Efficient Cooperative Communication", *IEEE VTC 2009, Anchorage, Alaska, USA*, 2009.
- [10] R. Deshmukh, and S. Ladhake, "Analysis of Various Puncturing Patterns and Code Rates: Turbo Code", *International Journal of Electronic Engineering Research* Volume 1, Number 2, pp.79-88, India, 2009.

AUTHOR PROFILE

Labib F.Gergis received the Bsc, Msc, and Ph.D from faculty of engineering, Mansoura University, Egypt, in 1980, 1990, and 2000, respectively. He is presently in Misr Academy for Engineering and Technology, Egypt. His areas of interest include digital communications, Coding, and Multiple Access.

Application of Fuzzy Composition Relation For DNA Sequence Classification

Amrita Priyam

Dept. of Computer Science and Engineering
Birla Institute of Technology
Ranchi, India.

B. M. Karan⁺, G. Sahoo⁺⁺

⁺Dept. of Electrical and Electronics Engineering
⁺⁺Dept. of Information Technology
Birla Institute of Technology
Ranchi, India

Abstract— *Abstract—* This paper presents a probabilistic approach for DNA sequence analysis. A DNA sequence consists of an arrangement of the four nucleotides A, C, T and G. There are various representation schemes for a DNA sequence. This paper uses a representation scheme in which the probability of a symbol depends only on the occurrence of the previous symbol. This type of model is defined by two parameters, a set of states Q , which emit symbols and a set of transitions between the states. Each transition has an associated transition probability, a_{ij} , which represents the conditional probability of going to state j in the next step, given that the current state is i . Further, the paper combines the different types of classification classes using a Fuzzy composition relation. Finally a log-odd ratio is used for deciding to which class the given sequence belongs to.

Keywords-component; Transition Probability, Fuzzy Composition Relation, Log-Odd ratio

I. INTRODUCTION

A DNA sequence is a succession of the letters A, C, T and G. The sequences are any combination of these letters. A physical or mathematical model of a system produces a sequence of symbols according to a certain probability associated with them. This is known as a stochastic process, that is, it is a mathematical model for a biological system which is governed by a set of probability measure. The occurrence of the letters can lead us to the further study of genetic disorder. There are various representation schemes for a DNA sequence. This paper uses a representation scheme in which the probability of a symbol depends only on the occurrence of the previous symbol and not on any other symbol before that. This type of model is defined by two parameters, a set of states Q , which emit symbols and a set of transitions between the states. Each transition has on associated transition probability, a_{ij} , which represents the conditional probability of going to state j from state i in the next step, given that the current state is i . Each

class has a set of transition probabilities associated with it. This transition probability is the measure of going from one state to another. Now, each class has a set of transition probability associated with it. We further group the similar classes and their respective transition probability is merged using a fuzzy composition relation. Finally a log-odd ratio is used for deciding to which class the given sequence belongs to.

II. DNA SEQUENCES

DNA sequence is a succession of letters representing the primary structure of a real or hypothetical DNA molecule or strand, with the capacity to carry information as described by the central dogma of molecular biology. There are 4 nucleotide bases (A – Adenine, C – Cytosine, G – Guanine, T – Thymine). DNA sequencing is the process of determining the exact order of the bases A, T, C and G in a piece of DNA [3]. In essence, the DNA is used as a template to generate a set of fragments that differ in length from each other by a single base. The fragments are then separated by size, and the bases at the end are identified, recreating the original sequence of the DNA[8][9]. The most commonly used method of sequencing DNA the dideoxy or chain termination method was developed by Fred Sanger in 1977 (for which he won his second Nobel Prize). The key to the method is the use of modified bases called dideoxy bases; when a piece of DNA is being replicated and a dideoxy base is incorporated into the new chain, it stops the replication reaction.

Most DNA sequencing is carried out using the chain termination method [4]. This involves the synthesis of new DNA strands on a single standard template and the random incorporation of chain-terminating nucleotide analogues. The chain termination method produces a set of DNA molecules differing in length by one nucleotide. The last base in each molecule can be identified by way of a unique label. Separation of these DNA molecules according to size places them in correct order to read off the sequence.

III. A PROBABILISTIC APPROACH FOR SEQUENCE REPRESENTATION

A DNA sequence is essentially represented as a string of four characters A, C, T, G and looks something like ACCTGACCTTACG. These strings can also be represented in terms of some probability measures and using these measures it can be depicted graphically as well. This graphical representation matches the Markov Hidden Model. A physical or mathematical model of a system produces a sequence of symbols according to a certain probability associated with them. This is known as a stochastic process [2]. There are different ways to use probabilities for depicting the DNA sequences. The diagrammatical representation can be shown as follows:

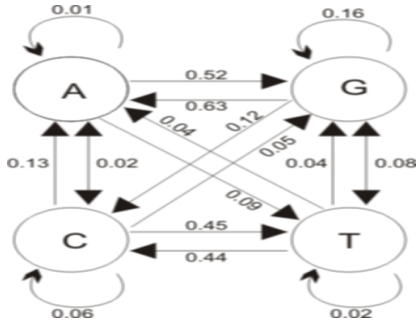


FIG 1: [The states of A, C, G and T.]

For example, the transition probability from state G to state T is 0.08, i.e,

$$P(x_i = T | x_{i-1} = G) = 0.08$$

In a given sequence x of length L , x_1, x_2, \dots, x_L , represent the nucleotides. The sequence starts at the first state x_1 , and makes successive transitions to x_2, x_3 and so on, till x_L . Using Markov property [6], the probability of x_L , depends on the value of only the previous state, x_{L-1} , not on the entire previous sequence. This characteristic is known as Markov property [5] and can be written as:

$$\begin{aligned} P(x) &= P(x_L | x_{L-1})P(x_{L-1} | x_{L-2}) \dots P(x_2 | x_1)P(x_1) \\ &= P(x_1) \prod_{i=2}^L P(x_i | x_{i-1}) \end{aligned} \quad (1)$$

In Equation (1) we need to specify $P(x_1)$, the probability of the starting state. For simplicity, we would like to model this as a transition too. This can be done by adding a begin state, denoted by 0, so that the starting state becomes $x_0=0$.

Now considering $a_{x_{i-1}x_i}$, the transition probability we can rewrite (1) as

$$P(x) = \prod_{i=1}^L a_{x_{i-1}x_i} \quad (2)$$

If there are n classes, then we calculate the probability of a sequence x being in all the classes. To overcome this drawback we use Fuzzy composition relation. That is, we divide the n classes into different groups based on their similarities. So, if out of n classes, m are similar then they are treated as one group and their individual transition probability tables are merged using the fuzzy composition relation. The remaining $(n - m)$ classes are similarly grouped. Lets say, if there are two classes R_1 and R_2 , the Fuzzy composition relation between R_1 and R_2 [6][7] can be written as follows:

$$R_1 \circ R_2 = \text{Max}(\text{Min}(R_1(x), R_2(y))) \quad (3)$$

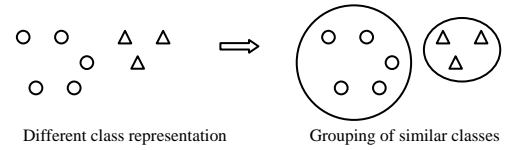


Fig 2: Grouping of similar classes

A table is then constructed representing the entire $(n - m)$ similar classes. From this table we compute the probability that a sequence x belongs to a given group using the following equation:

$$\log \frac{P(x|+)}{P(x|-)} = \sum_{i=1}^L \log \frac{a_{i-1}^+ x_i}{a_{i-1}^- x_i} \quad (4)$$

Here “+” represents transition probability of the sequence belonging to one of the classes using fuzzy composition relation and “-” represents the transition probability of the same for another class [1].

If this ratio is greater than zero then we can say that the sequence x is from the first class else from the other one.

An Example:

Let us consider an example for applying this classification method. We have taken into consideration the Swine flu data.[11] The different categories of the Swine flu data are shown as R_1, R_2 and R_3 .

R_1, R_2 and R_3 shows the Transition Probability of Type 1, Type 2 and Type 3 varieties of Avian Flu.

$$R_1 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.13 & 0.06 & 0.09 & 0.08 \\ 0.09 & 0.04 & 0.05 & 0.02 \\ 0.06 & 0.05 & 0.05 & 0.07 \\ 0.08 & 0.04 & 0.05 & 0.06 \end{bmatrix} \end{matrix}$$

$$R_2 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.13 & 0.06 & 0.09 & 0.08 \\ 0.08 & 0.04 & 0.04 & 0.02 \\ 0.06 & 0.05 & 0.06 & 0.07 \\ 0.08 & 0.04 & 0.04 & 0.06 \end{bmatrix} \end{matrix}$$

$$R_3 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.08 & 0.05 & 0.08 & 0.08 \\ 0.08 & 0.03 & 0.06 & 0.02 \\ 0.05 & 0.06 & 0.06 & 0.09 \\ 0.08 & 0.06 & 0.04 & 0.07 \end{bmatrix} \end{matrix}$$

Using the Fuzzy composition relation technique on R_1 and R_2 and then the result of the application with relation R_3 we get the final table for the Swine Flu class as:

$$\begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.08 & 0.06 & 0.08 & 0.09 \\ 0.08 & 0.06 & 0.08 & 0.09 \\ 0.07 & 0.06 & 0.07 & 0.07 \\ 0.08 & 0.06 & 0.08 & 0.08 \end{bmatrix} \end{matrix}$$

Similarly, we can repeat the same procedure for another class Staphylococcus. X_1 , X_2 and X_3 shows the Transition Probability of Type 1, Type 2 and Type 3 varieties of Staphylococcus.

$$X_1 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.08 & 0.05 & 0.07 & 0.06 \\ 0.06 & 0.05 & 0.06 & 0.08 \\ 0.04 & 0.06 & 0.08 & 0.07 \\ 0.08 & 0.09 & 0.03 & 0.05 \end{bmatrix} \end{matrix}$$

A C T G

$$X_2 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.09 & 0.06 & 0.06 & 0.06 \\ 0.07 & 0.05 & 0.06 & 0.07 \\ 0.04 & 0.05 & 0.07 & 0.07 \\ 0.07 & 0.09 & 0.04 & 0.05 \end{bmatrix} \end{matrix}$$

$$X_3 = \begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.07 & 0.05 & 0.07 & 0.04 \\ 0.05 & 0.06 & 0.06 & 0.08 \\ 0.03 & 0.06 & 0.04 & 0.09 \\ 0.06 & 0.09 & 0.06 & 0.07 \end{bmatrix} \end{matrix}$$

Applying Fuzzy composition relation to these tables we get the final table as

$$\begin{matrix} & \begin{matrix} A & C & T & G \end{matrix} \\ \begin{matrix} A \\ C \\ T \\ G \end{matrix} & \begin{bmatrix} 0.07 & 0.06 & 0.05 & 0.06 \\ 0.08 & 0.09 & 0.07 & 0.06 \\ 0.04 & 0.03 & 0.07 & 0.06 \\ 0.05 & 0.08 & 0.04 & 0.02 \end{bmatrix} \end{matrix}$$

Suppose we are given the sequence x which is to be classified as either falling into any of the given classes and say $x = CGCG$

From the final fuzzy composition table the log odds ratio of this sequence is:

$$\log \frac{0.09}{0.08} + \log \frac{0.06}{0.07} + \log \frac{0.09}{0.08} = 0.032270$$

Now, since this ratio is greater than 0, we can conclude that the input sequence x belongs to the class Avian Flu. If further classification on the data is needed we will then consult the individual transition probabilities for all the three types.

CONCLUSION

In this paper we have used a probabilistic function for the Markov Property. We have applied this for probabilistic determination in the case of Avian flu virus and Staphylococcus. The paper also presented a way for identifying particular classes of genes or proteins. A given input sequence can belong to either of the given classes. By using a transition probability measure, one had to determine a value for each class even though they were similar. The paper presented a scheme such that the similar classes were merged by using the fuzzy composition relation and now instead of calculating each individual probability measure, one measure

is sufficient to depict all the similar classes. This measure is further used in the log odds ratio to finally predict the class of the input sequence.

REFERENCES

- [1] Anna Loanova, "Introduction to (log) odds ratio statistics and Methodology", University of Groningen, 2008.
- [2] C. E. Shanon, "A Mathematical Theory of Communciation", The Bell System Technical Journal, Vol. 27, pp. 379 – 423, 623 – 656.
- [3] D. W. Mount, " Bioinformatics, Sequence and Genome Analysis", 2nd edition, CSHL Press, (2004). (3)
- [4] Durbin, Eddy, Krogh, Mitchison, "Biological Sequence Analysis", Cambridge University Press, 1998.
- [5] L. R Rabiner, "A tutorial on Hidden Markov Models and selected Application in speech recognition" -Proceeding of the IEEE, Vol.77, No.2 Feb.1989.
- [6] Lee, Kwang Hyung, "First course on Fuzzy Theory and Applications", Advances in Soft Computing, Vol. 27, Springer, 2005.
- [7] Michael Hanss , "Applied Fuzzy Arithmetic : An Introduction with Engineering Applications", Springer, 2005.
- [8] T. Dewey and M. Herzel, "Application of Information Theory to Biology", Pacific Symposium on Biocomputing, 5:597 – 598 (2000).
- [9] W. J. Ewens, G. R. Grant, "Statistical Methods in Bioinformatics: An Introduction", Vol. 13, 2nd edition, Springer.
- [10] Y. Ephraim, L. R. Rabiner, "On the Relations Between Modeling Approaches for Speech Recognition", IEEE transactions on Information Theory, vol. 36, no. 2, March 1990.
- [11] A.Priyam, B.M.Karan, G.Sahoo, "A Probabilistic Model For Sequence Analysis", Inter national Journal of Computer Science and Information security vol7 No.1, (2010) 244-247.

Data Security in Mobile Ad Hoc Networks using Genetic Based Biometrics

B. Shanthini, Research Scholar
CSE Department
Anna University
Chennai, India

S. Swamynathan, Assistant Professor
CSE Department
Anna University
Chennai, India

Abstract— A mobile ad hoc network (MANET) is a self configuring, dynamic, multi hop radio network without any fixed infrastructure. MANETs are collections of wireless mobile devices with restricted broadcast range and resources and communication is achieved by relaying data along appropriate routes that are dynamically discovered and maintained through collaboration between the nodes. The main challenge in the design of such networks is how to prevent the attacks against data such as unauthorized data modification, impersonation etc. Biometrics provides possible solutions for this security problem in MANET since it has the direct connection with user identity and needs little user interruption. So, researchers have been investigating ways to use biometric features of the user rather than memorable password or passphrase, in an attempt to produce tough and repeatable cryptographic keys. In this paper such a security system based on Biometrics and Genetic algorithm which is providing data security in MANET is presented.

Keywords— Mobile Ad hoc Networks, Data Security, Biometrics, Genetic Algorithm.

I. INTRODUCTION

Mobile ad hoc networks are seen as autonomous that can be quickly formed, on demand, for specific tasks and mission support. Communication generally happens through wireless links, in which nodes within a radio range communicate and coordinate to create a virtual and temporary communication infrastructure for data routing and data transmission. MANET can operate in isolation or in coordination with a wired network through a gateway node participating in both networks. This flexibility along with their self-organizing capabilities, are some of their biggest strengths, as well as their biggest security weaknesses.

The applications of MANET include the foremost situations such as emergency/crisis management, military, healthcare, disaster relief operations and intelligent transportation systems. So message security plays a vital role in data transmission in MANET. However, because of the absence of an established infrastructure or centralized administration, implementation of hard-cryptographic algorithms is a challenging prospect. So, in this paper, we present a novel security method using genetic based biometric cryptography for message security in mobile ad hoc networks.

A. Security challenges in MANET

Wireless ad hoc networks are vulnerable to various attacks [1]. Adversaries may attempt passive and active attacks to gain unauthorized access to classified information, modify the information, delete the information or disrupt the information flow. The best way to protect data information in a most fine-granular way is by providing security at the application layer. It is highly desirable to handle data confidentiality and integrity in application layer, since this is the easiest way to protect data from altering, fabrication and compromise. With the rapid evolution of wireless technology the reliance of ad hoc networks to carry mission critical information is rapidly growing. This is especially important in a military scenario where strategic and tactical information is sent. Therefore the ability to achieve a highly secure authentication is becoming more critical.

Numerous countermeasures such as strong authentication, encrypting and decrypting the messages using traditional cryptographic algorithms and redundant transmission can be used to tackle these attacks. Even though these traditional approaches play an important role in achieving confidentiality, integrity, authentication and non-repudiation, these are not sufficient for more sensitive and mission-critical applications and they can address only a subset of the threats. Moreover, MANETs [2] cannot support complex computations or high communication over head due to the limited memory and limited computation power of mobile nodes.

B. Necessity of Biometrics Security

For mission-critical applications such as a military application may have higher requirements regarding data or information security. In such a scenario, we may design the security system combining both biometrics and cryptography. Biometric based security scheme overcome the limitations of traditional security solutions. Biometrics refers to the methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits like fingerprints, iris, retina scans, hand, face, ear geometry, hand vein, nail bed, DNA, palm print, signature, voice, keystroke or mouse dynamics, and gait analysis etc.

Biometric technologies have confirmed its importance in the fields such as security, access control and monitoring applications. The tradeoffs among these biometric technologies really depend on the application and security level involved. The best biometric technology [3] that can easily be deployable in ad hoc networks is fingerprint recognition. Fingerprints have been successfully used in civilian identification for years because of their unchangeability during the human life time and uniqueness of each individual. As biometrics can't be borrowed, stolen, or forgotten, and forging is practically impossible, it has been presented as a natural identity tool that offers greater security and convenience than traditional methods of personal recognition.

Even though biometric has advantages, it also raises many security and privacy concerns as given below:

- i. Biometric is authentic but not secret.
- ii. Biometric cannot be revoked or cancelled.
- iii. If a biometric is lost once, it is compromised forever.
- iv. Cross-matching can be used to track individuals without their consent.

To overcome these disadvantages, instead of using the original biometric, a set of features are taken from it and transformed using genetic algorithm. If a biometric is compromised, it can be simply reenrolled using another feature set and another genetic operation, thus providing revocability and the privacy of the biometric is preserved.

C. Genetic Algorithms

Genetic algorithms [4] are a family of computational models inspired by natural evolution. They belong to the field of evolutionary computation and are based on three main operators: Selection selects the fittest individuals, called parents that contribute to the reproduction of the population at the next generation, Crossover combines two parents to form children for the next generation and Mutation applies random changes to individual parents to form children. Two-point crossover operator is used here which has the ability to generate, promote, and juxtapose building blocks to form the optimal strings.

This paper is organized into 4 sections. Section 1 introduces the background and initiatives of the research. It also discusses the challenges of message security, the necessity of biometric security in MANET and Genetic algorithms. Section 2 explains the related research works that has been done to provide security in MANET. Section 3 proposes a new security scheme for MANET which combines genetic algorithm and biometrics. Section 4 contains conclusion and suggestions for future research.

II. RELATED WORK

A few research works that has been done for data security in MANET, the various approaches of biometric security and Genetic algorithms in security are briefly presented.

Qinghan Xiao [5] introduced a new strategy for authentication of mobile users. Each user has a profile which contains all the information of the ID holders. The group leader also maintains the biometric templates of the group members. Instead of a central authentication server, the group leaders act as distributed authenticators. Each group has a shared cryptographic key which is used for cryptographic communication within the group. The proposed approach is designed for high security small group coalition operations and may not be suitable for enterprise usage.

Jie Liu et al. [6] proposed an optimal biometric-based continuous authentication scheme in MANET which distinguished two classes of authentications: user-to-device and device-to-network. This model focused on the user-to-device class and it can optimally control whether or not to perform authentication as well as which biometrics to use to minimize the usage of system resources.

B Ananda Krishna et al. [7] depicted a model which used multiple algorithms for encryption and decryption. Each time a data packet is sent to the application layer it is encrypted using one of these randomly selected algorithms. When responses are analyzed they give a random pattern and difficult to know neither algorithms nor keys. The proposed scheme worked well for heavily loaded networks with high mobility.

Zarza L et al. [8] explained the context of the study of Genetic Algorithms as an aiding tool for generating and optimizing security protocols. This paper explains how security protocols can be represented as binary strings, how GA tools are used to define genome interpretation in optimization problems.

B. Shanthini et al. [9] explained Cancelable Biometric-Based Security System (CBBSS), where cancelable biometrics is used for data security in mobile ad hoc networks. Fingerprint feature of the receiver is coupled with the tokenized random data by using inner-product algorithm and this product is discretized based on a threshold to produce a set of private binary code which is acting as a cryptographic key in this system.

A. Jagadeesan et al. [10], proposed an efficient approach based on multimodal biometrics (Iris and fingerprint) for generating a secure cryptographic key, where the security is further enhanced with the difficulty of factoring large numbers. At first, the features, minutiae points and texture properties are extracted from the fingerprint and iris images respectively. Then, the extracted features are fused at the feature level to obtain the multi-biometric template. Finally, a multi-biometric template is used for generating a 256-bit cryptographic key.

III. PROPOSED WORK

In this proposed Genetic-Based Biometric Security System (GBBSS), a genetic two-point crossover operator is applied on biometric feature set and is used for data security in mobile ad hoc networks. The main objective of the proposed security scheme is to improvise the existing data security approaches for MANET to suit technology enhancements and to study the network performance.

A. Generation of Genetic-Based Biometric Key

In this model all the group members maintain the biometric templates of the other group members. Suppose a member wants to send a message to any other member, the receiver's fingerprint is divided into slices and feature set taken from the slices is undergone a genetic two-point crossover operation and the result is the cryptographic key in this system. Generation of cryptographic key is shown in figure 1.

Fingerprint

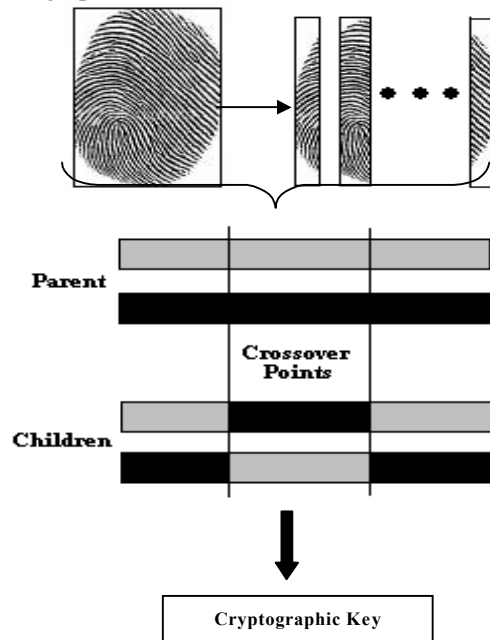


Figure 1. Generation of cryptographic key from the finger print features.

The same key is generated by the receiver by using his biometric and the same sort of cross over operations and is used for decryption.

Example:

Parent	01011100	1010000011111010	00110101
	00110011	1111000011110000	11001100
After Crossover			
Children	01011100	1111000011110000	00110101
	00110011	1010000011111010	11001100

If this biometric based key is compromised a new one can be issued by using a different set of features and different cross over operation and the compromised one is rendered completely useless. It can also be an application specific that is different sets of fingerprint features can be used with different cross over operations to generate respective cryptographic key for different applications.

B. Securing the Data

Data is secured by applying this cryptographic key to encrypt the actual message using a simple cryptographic algorithm say Fiestel algorithm. The encryption and decryption processes are specified by the formulae:

$$C = E_{KR} (P) \text{ and } P = D_{KR} (C)$$

where P – Plain Text
 C - Cipher Text
 KR - Key created by Receiver's Biometric
 E - Encryption Algorithm
 D - Decryption Algorithm

In Fiestel algorithm, a block of size N is divided into two halves, of length $N/2$, the left half called XL and right half called XR . The output of the i th round is determined from the output of the $(i-1)$ th round. The same key is used for all iterations without generating sub keys. Also the number of iterations performed is reduced to show that security can be achieved by using simple algorithm. For example if the plaintext is of 512 bytes, then encryption is performed for every 64 bits and the process is repeated until all 512 bytes are encrypted. Fiestel structure is given in figure 2. [1].

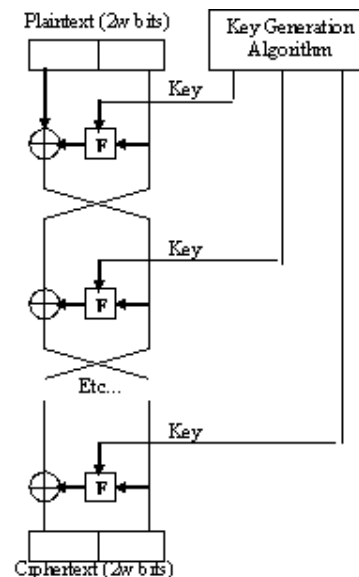


Figure 2. Fiestel Algorithm

Algorithm for Encryption:

1. Divide the plaintext into two blocks of size, 32 bytes, XL and XR
2. For $I = 1$ to 32
Do $XL = XL \text{ XOR Key}$
 $XR = F(XL) \text{ XOR XR}$
Swap XL, XR
Join XL, XR
3. Repeat step 2 until the entire plaintext is encrypted

Algorithm for Decryption:

Do the reverse operation of Encryption process.

C. Implementation of GBBSS in MANET

The proposed scheme can be implemented over any unicast routing protocols like DSR or AODV which discover routes as and when necessary and the routes are maintained just as long as necessary. A typical MANET is shown in figure 3.

Suppose User A wants to send the message to User C, after the forward and reverse paths are set up by the route discovery method, the data will be sent through that path to the destination C. Before sending the data through that path, the data will be encrypted by Fiestel algorithm using the genetic based biometric key. Once the cipher text is received by the receiver, the cipher text is decrypted by using the same key.

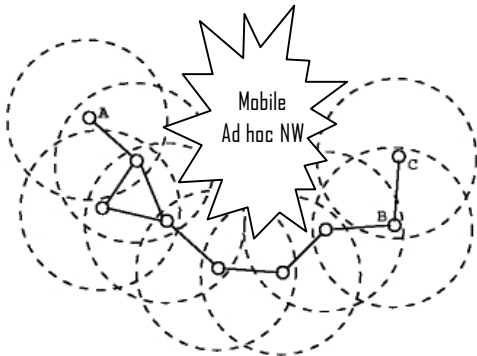


Figure 3. MANET Structure

D. The security functions of the proposed system

- **Confidentiality:** The privacy of the message is protected by this scheme. Suppose if the attacker wants to derive the original message from the encrypted text, he needs the cryptographic key. The key can be obtained only by using the biometric of the receiver. Furthermore the biometric is not used as such instead a cancelable version is used. So, it is computationally infeasible to get the key.
- **Authentication:** In our proposed scheme, the members of the ad hoc group can authenticate each other through their biometric. If the receiver wants to verify whether the message is coming from the genuine sender, the message can be encrypted by using the sender's biometric and the receiver can use the same biometric to decrypt the message. These processes can be specified by the following formulae:

$$C = E_{K_S} (P) \text{ and } P = D_{K_S} (C)$$

where K_S is the Key created by Sender's Biometric.

- **Integrity:** In our proposed scheme, the recipient can verify whether the received message is the original one that was sent by the sender. If the attacker changes the cipher text, the original plain text can not be generated after decrypting with the key created by using receiver's biometric. By the property of one-way hash function, it is computationally infeasible for the attacker to modify the cipher text.
- **Man-in-the-middle attack:** An attacker sits between the sender and the receiver and sniffs any information being sent between two ends is called man in the middle attack.

Even though the attacker can get the cipher text he cannot view the original message since it is secured using genetic based biometric cryptography.

E. Security Analysis

This section reports the analysis of the security parameters like time taken for key generation, encryption and decryption for various algorithms like 3DES192, AES128, AES256 and GBBSS64 in an ad hoc network environment. The graphs shown in figure 4 and figure 5 are generated by using the values given in the following table 1:

Encryption Algorithm	Parameters			
	Key Size	Time taken for Key Generation	Time taken for Encryption	Time taken for Decryption
3DES192	192	0.08 ms	0.08 ms	0.07 ms
AES-128	128	0.13 ms	0.1 ms	0.1 ms
AES-256	256	0.13 ms	0.12 ms	0.11 ms
GBBSS-64	64	0.06 ms	0.04 ms	0.02 ms

Table 1: Key size and Timing measurements for various algorithms

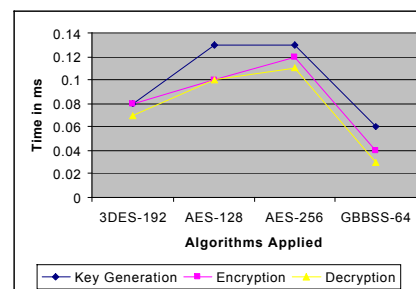


Figure 4. Timing measurements for various algorithms

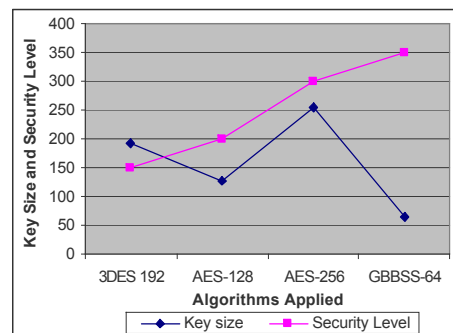


Figure 5. Key Size and Security Levels for various algorithms

From the above charts we can understand that our proposed GBBSS achieves relatively high performance in terms of less overhead and high security level. Since the key size is very small compared to the other algorithms, the time taken to generate the key, time taken to encrypt and decrypt are also less.

IV. CONCLUSION AND FUTURE WORK

Although MANET is a very promising technology, challenges are slowing its development and deployment. Traditional security mechanisms are not sufficient for the nodes roaming in a hostile environment with relatively poor physical protection. Therefore to strengthen the encryption algorithm and key, first the advantages of biometric and genetic algorithms are taken into our system. Secondly, security should be achieved by using simple algorithms that involve small inherent delays rather than complex algorithms which occupy considerable memory and delay. Finally, ad hoc network may consist of thousands of nodes. So, security mechanisms should be scalable to handle such a large network.

The method presented in this paper remains as a preliminary approach to realize biometric security in ad hoc networks which needs high security. This approach can be used in very critical, crucial and vital applications where data security is very important and members who have accessed that data is limited in number like military officers at war-field, scientists in a confidential conference, officers in the intelligent buildings etc. There are many security problems still persist in these types of ad-hoc networks and as a future work, this paper can be extended to solve those problems with different biometrics and also with multimodal biometrics.

REFERENCES

- [1] Stallings W, "Cryptography and Network Security-Principles and Practices", 3rd Edition, Pearson Education, 2004.
- [2] Animesh K. Trivedi, Rajan Arora, Rishi Kapoor, Sudip Sanyal, Ajith Abraham, Sugata Sanyal, "Mobile Ad Hoc Network Security Vulnerabilities", IGI Global, 2009.
- [3] Maltoni D. Maio, Jain A. K. and Prabhakar S, "Handbook of Fingerprint Recognition", Springer Verlag, 2003.
- [4] Fessi B A, Ben Abdallah, S, Hamdi Mand Boudriga, "A new genetic algorithm approach for intrusion response system in computer networks", IEEE Symposium on Computers and Communications, pp. 342-347, 2009.
- [5] Qinghan Xiao, "A Biometric Authentication Approach for High Security Ad hoc Networks", Proceedings of IEEE Workshop on Information Assistance, pp. 250-256, June 2004.
- [6] Jie Liu, F. Richard Yu, Chung-Horng Lung and Helen Tang, "Optimal Biometric-Based Continuous Authentication in Mobile Ad hoc Networks", Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pp. 76-81, 2007.
- [7] B Ananda Krishna, S Radha and K Chenna Kesava Reddy, "Data Security in Ad hoc Networks using Randomization of Cryptographic Algorithms", Journal of Applied Sciences, pp. 4007-4012, 2007.
- [8] Zarza L., Pegueroles J and Soriano M "Interpretation of Binary Strings as Security Protocols for their Evolution by means of Genetic Algorithms", International Conference on Database and Expert Systems Applications, pp. 708-712, 2007.
- [9] B. Shanthini and S. Swamynathan "A Cancelable Biometric-Based Security System for Mobile Ad Hoc Networks", International Conference on Computer Technology (ICONCT 09), pp. 179-184, December, 2009.
- [10] A. Jagadeesan, T. Thillaikkarasi and K. Duraiswamy, "Cryptographic Key Generation from Multiple Biometric Modalities: Fusing Minutiae with Iris Feature", International Journal of Computer Applications, Vol. 2, No.6, pp. 0975-8887, June 2010.



B. Shanthini is a research scholar in Anna University, Chennai, India. She received her Bachelor's degree in C.S.E. from M.K.University, Madurai and Master's degree in C.S.E. from M.S. University, Tirunelveli. Her research interests include Network Security, Web Security, Wireless Communication, Biometrics and Cloud Computing.



Dr. S. Swamynathan is an Assistant Professor of Computer Science and Engineering at Anna University Chennai, India. He received his Master's in Computer Science and Engineering and Doctorate in Reactive Web Services from Anna University, Chennai. His research interests include Web Service, Security, Web Mining and Automated Workflow Systems.

Effective Multi-Stage Clustering for Inter- and Intra-Cluster Homogeneity

Sunita M. Karad[†]

Assistant Professor of Computer Engineering,
MIT, Pune, INDIA

V.M.Wadhai^{††}

Professor and Dean of Research, MITSOT,
MAE, Pune, INDIA

M.U.Kharat^{†††}

Principle of Pankaj Laddhad IT,
Yelgaon, Buldhana, INDIA

Prasad S.Halgaonkar^{††††}

Faculty of Computer Engineering,
MITCOE, Pune, INDIA

Dipti D. Patil^{†††††}

Assistant Professor of Computer Engineering,
MITCOE, Pune, INDIA

Abstract - A new algorithm for clustering high-dimensional categorical data is proposed and implemented by us. This algorithm is based on a two-phase iterative procedure and is parameter-free and fully-automatic. Cluster assignments are given in the first phase, and a new cluster is added to the partition by identifying and splitting a low-quality cluster. Optimization of clusters is carried out in the second phase. This algorithm is based on quality of cluster in terms of homogeneity. Suitable notion of cluster homogeneity can be defined in the context of high-dimensional categorical data, from which an effective instance of the proposed clustering scheme immediately follows. Experiment is carried out on real data; this innovative approach leads to better inter- and intra-homogeneity of the clusters obtained.

Index Terms - Clustering, high-dimensional categorical data, information search and retrieval.

I. INTRODUCTION

Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attributes (dimensions) [1] [2]. Clustering techniques have been studied extensively in statistics, pattern recognition, and machine learning. Recent work in the database community includes CLARANS, BIRCH, and DBSCAN. Clustering is an unsupervised classification technique. A set of unlabeled objects are grouped into meaningful clusters, such that the groups formed are homogeneous and neatly separated. Challenges for clustering categorical data are: 1) Lack of ordering of the domains of the individual attributes. 2) Scalability to high dimensional data in terms of

effectiveness and efficiency. High-dimensional categorical data such as market-basket has records containing large number of attributes. 3) Dependency on parameters. Setting of many input parameters is required for many of the clustering techniques which lead to many critical aspects.

Parameters are useful in many ways. Parameters support requirements such as efficiency, scalability, and flexibility. For proper tuning of parameters a lot of effort is required. As number of parameters increases, the problem of parameter tuning also increases. Algorithm should have as less parameters as possible. If the algorithm is automatic it helps to find accurate clusters. An automatic approach technique searches huge amounts of high-dimensional data such that it is effective and rapid which is not possible for human expert. A parameter free approach is based on decision tree learning, which is implemented by top-down divide-and-conquer strategies. The above mentioned problems have been tackled separately, and specific approaches are proposed in the literature, which does not fit the whole framework. The main objective of this paper is to face the three issues in a unified framework. We look forward to an algorithmic technique that is capable of automatically detecting the underlying interesting structure (when available) on high-dimensional categorical data.

We present Two Phase Clustering (MPC), a new approach to clustering high-dimensional categorical data that scales to processing large volumes of such data in terms of both effectiveness and efficiency. Given an initial data set, it searches for a partition, which improves the overall purity. The algorithm is not dependent on any data-specific parameter (such as the number of clusters or occurrence thresholds for frequent attribute values). It is intentionally left parametric to

the notion of purity, which allows for adopting the quality criterion that best meets the goal of clustering. Section-2 reviews some of the related work carried out on transactional data, high dimensional data and high dimensional categorical data. Section-3 provides background information on the clustering of high dimensional categorical data (MPC algorithm). Section-4 describes implementation results of MPC algorithm. Section-5 concludes the paper and draws direction to future work.

II. RELATED WORK

In current literature, many approaches are given for clustering categorical data. Most of these techniques suffer from two main limitations, 1) their dependency on a set of parameters whose proper tuning is required and 2) their lack of scalability to high dimensional data. Most of the approaches are unable to deal with the above features and in giving a good strategy for tuning the parameters.

Many distance-based clustering algorithms [3] are proposed for transactional data. But traditional clustering techniques have the curse of dimensionality and the sparseness issue when dealing with very high-dimensional data such as market-basket data or Web sessions. For example, the K-Means algorithm has been adopted by replacing the cluster mean with the more robust notion of cluster medoid (that is, the object within the cluster with the minimal distance from the other points) or the attribute mode [4]. However, the proposed extensions are inadequate for large values of m : Gozzi et al. [5] describe such inadequacies in detail and propose further extensions to the K-Means scheme, which fit transactional data. Unfortunately, this approach reveals to be parameter laden. When the number of dimensions is high, distance-based algorithms do not perform well. Indeed, several irrelevant attributes might distort the dissimilarity between tuples. Although standard dimension reduction techniques [6] can be used for detecting the relevant dimensions, these can be different for different clusters, thus invalidating such a preprocessing task. Several clustering techniques have been proposed, which identify clusters in subspaces of maximum dimensionality (see [7] for a survey). Though most of these approaches were defined for numerical data, some recent work [8] considers subspace clustering for categorical data.

A different point of view about (dis)similarity is provided by the ROCK algorithm [9]. The core of the approach is an agglomerative hierarchical clustering procedure based on the concepts of neighbors and links. For a given tuple x , a tuple y is a neighbor of x if the Jaccard similarity $J(x, y)$ between them exceeds a prespecified threshold Θ . The algorithm starts by assigning each tuple to a singleton cluster and merges clusters on the basis of the number of neighbors (links) that they share until the desired number of clusters is reached. ROCK is robust to high-dimensional data. However, the dependency of the algorithm to the parameter Θ makes proper tuning difficult.

Categorical data clusters are considered as dense regions within the data set. The density is related to the frequency of particular groups of attribute values. The higher

the frequency of such groups the stronger the clustering. Preprocessing the data set is carried by extracting relevant features (frequent patterns) and discovering clusters on the basis of these features. There are several approaches accounting for frequencies. As an example, Yang et al. [10] propose an approach based on histograms: The goodness of a cluster is higher if the average frequency of an item is high, as compared to the number of items appearing within a transaction. The algorithm is particularly suitable for large high-dimensional databases, but it is sensitive to a user defined parameter (the repulsion factor), which weights the importance of the compactness/sparseness of a cluster. Other approaches [11], [12], [13] extend the computation of frequencies to frequent patterns in the underlying data set. In particular, each transaction is seen as a relation over some sets of items, and a hyper-graph model is used for representing these relations. Hyper-graph partitioning algorithms can hence be used for obtaining item/transaction clusters.

The CLICKS algorithm proposed in [14] encodes a data set into a weighted graph structure $G(N, E)$, where the individual attribute values correspond to weighted vertices in N , and two nodes are connected by an edge if there is a tuple where the corresponding attribute values co-occur. The algorithm starts from the observation that clusters correspond to dense (that is, with frequency higher than a user-specified threshold) maximal k -partite cliques and proceeds by enumerating all maximal k -partite cliques and checking their frequency. A crucial step is the computation of strongly connected components, that is, pairs of attribute values whose co-occurrence is above the specified threshold. For large values of m (or, more generally, when the number of dimensions or the cardinality of each dimension is high), this is an expensive task, which invalidates the efficiency of the approaches. In addition, technique depends upon a set of parameters, whose tuning can be problematic in practical cases.

Categorical clustering can be tackled by using information-theoretic principles and the notion of entropy to measure closeness between objects. The basic intuition is that groups of similar objects have lower entropy than those of dissimilar ones. The COOLCAT algorithm [15] proposes a scheme where data objects are processed incrementally, and a suitable cluster is chosen for each tuple such that at each step, the entropy of the resulting clustering is minimized. The *scalable Information Bottleneck* (LIMBO) algorithm [16] also exploits a notion of entropy to catch the similarity between objects and defines a clustering procedure that minimizes the information loss. The algorithm builds a Distributional Cluster Features (DCF) tree to summarize the data in k clusters, where each node contains statistics on a subset of tuples. Then, given a set of k clusters and their corresponding DCFs, a scan over the data set is performed to assign each tuple to the cluster exhibiting the closest DCF. The generation of the DCF tree is parametric to a user-defined branching factor and an upper bound on the distance between a leaf and a tuple.

Li and Ma [17] propose an iterative procedure that is aimed at finding the optimal data partition that minimizes an

entropy-based criterion. Initially, all tuples reside within a single cluster. Then, a Monte Carlo process is exploited to randomly pick a tuple and assign it to another cluster as a trial step aimed at decreasing the entropy criterion. Updates are retained whenever entropy diminishes. The overall process is iterated until there are no more changes in cluster assignments. Interestingly, the entropy-based criterion proposed here can be derived in the formal framework of probabilistic clustering models. Indeed, appropriate probabilistic models, namely, multinomial [18] and multivariate Bernoulli [19], have been proposed and shown to be effective. The classical Expectation-Maximization framework [20], equipped with any of these models, reveals to be particularly suitable for dealing with transactional data [21], [22], being scalable both in n and in m . The correct estimation of an appropriate number of mixtures, as well as a proper initialization of all the model parameters, is problematic here.

The problem of estimating the proper number of clusters in the data has been widely studied in the literature. Many existing methods focus on the computation of costly statistics based on the within-cluster dispersion [23] or on cross-validation procedures for selecting the best model [24], [25]. The latter requires an extra computational cost due to a repeated estimation and evaluation of a predefined number of models. More efficient schemes have been devised in [26], [27]. Starting from an initial partition containing a single cluster, the approaches iteratively apply the K-Means algorithm (with $k = 2$) to each cluster so far discovered. The decision on whether to switch the original cluster with the newly generated sub-clusters is based on a quality criterion, for example, the Bayesian Information Criterion [26], which mediates between the likelihood of the data and the model complexity, or the improvement in the rate of distortion (the variance in the data) of the sub-clusters with respect to the original cluster [27]. The exploitation of the K-Means scheme makes the algorithm specific to low-dimensional numerical data, and proper tuning to high-dimensional categorical data is problematic.

Automatic approaches that adopt the top-down induction of decision trees are proposed in [28], [29], [30]. The approaches differ in the quality criterion adopted, for example reduction in entropy [28], [29] or distance among the prototypes of the resulting clusters [29]. All of these approaches have some of the drawbacks. The scalability on high-dimensional data is poor. Some of the literature that focused on high dimensional categorical data is available in [31], [32].

III. The MPC Algorithm

The key idea of Two Phase Clustering (MPC) algorithm is to develop a clustering procedure, which has the general sketch of a top-down decision tree learning algorithm. First, start from an initial partition which contains single cluster (the whole data set) and then continuously try to split a cluster within the partition into two sub-clusters. If the sub-clusters have a higher homogeneity in the partition than the original cluster, the original is removed. The sub-clusters obtained by

splitting are added to the partition. Split the clusters on the basis of their homogeneity. A function $Quality(C)$ measures the degree of homogeneity of a cluster C . Clusters with high intra-homogeneity exhibit high values of Quality.

Let M be set of Boolean attributes such that $M = \{a_1, \dots, a_m\}$ and a data set $D = \{x_1, x_2, \dots, x_n\}$ of tuples which is defined on M . A $a \in M$ is denoted as an item, and a tuple $x \in D$ as a transaction x . Data sets containing transactions are denoted as transactional data, which is a special case of high-dimensional categorical data. A cluster is a set S which is a subset of D . The size of S is denoted by n_S , and the size of $M_S = \{a | a \in x, x \in S\}$ is denoted by m_S . A partitioning problem is to divide the original collection of data D into a set $P = \{C_1, \dots, C_k\}$ where each clusters C_j are nonempty. Each cluster contains a group of homogeneous transactions. Clusters where transactions have several items have higher homogeneity than other subsets where transactions have few items. A cluster of transactional data is a set of tuples where few items occur with higher frequency than somewhere else.

Our approach to clustering starts from the analysis of the analogies between a clustering problem and a classification problem. In both cases, a model is evaluated on a given data set, and the evaluation is positive when the application of the model locates fragments of the data exhibiting high homogeneity. A simple rather intuitive and parameter-free approach to classification is based on decision tree learning, which is often implemented through top-down divide and conquers strategies. Here, starting from an initial root node (representing the whole data set), iteratively, each data set within a node is split into two or more subsets, which define new sub-nodes of the original node. The criterion upon which a data set is split (and, consequently, a node is expanded) is based on a quality criterion: choosing the best “discriminating” attribute (that is, the attribute producing partitions with the highest homogeneity) and partitioning the data set on the basis of such attribute. The concept of homogeneity has found several different explanations (for example, in terms of entropy or variance) and, in general, is related to the different frequencies of the possible labels of a target class.

The general schema of the MPC algorithm is specified in Fig. 1. The algorithm starts with a partition having a single cluster i.e whole data set (line 1). The central part of the algorithm is the body of the loop between lines 2 and 15. Within the loop, an effort is made to generate a new cluster by 1) choosing a candidate node to split (line 4), 2) splitting the candidate cluster into two sub-clusters (line 5), and (line 3) calculating whether the splitting allows a new partition with better quality than the original partition (lines 6–13). If this is true, the loop can be stopped (line 10), and the partition is updated by replacing the original cluster with the new sub-clusters (line 8). Otherwise, the sub-clusters are discarded, and a new cluster is taken for splitting.

The generation of a new cluster calls STABILIZE-CLUSTERS in line 9, improves the overall quality by trying relocations among the clusters. Clusters at line 4 are taken in increasing order of quality.

a. *Splitting a Cluster*

A splitting procedure gives a major improvement in the quality of the partition. Choose the attribute that gives the highest improvement in the quality of the partition.

```

GENERATE-CLUSTERS( $D$ )
Input: A set  $D = \{x_1, \dots, x_N\}$  of transactions;
Output: A partition  $P = \{C_1, \dots, C_k\}$  of clusters;
1. Let initially  $P = \{D\}$ ;
2. repeat
3.   Generate a new cluster  $C$  initially empty;
4.   for each cluster  $C_i \in P$  do
5.     PARTITION-CLUSTERS( $C_i, C$ );
6.      $P' \leftarrow P \cup \{C\}$ ;
7.     if  $Quality(P) < Quality(P')$  then
8.        $P \leftarrow P'$ ;
9.     STABILIZE-CLUSTERS( $P$ );
10.    break
11.  else
12.    Restore all  $x_j \in C$  into  $C_i$ ;
13.  end if
14. end for
15. until no further cluster  $C$  can be generated
  
```

Figure 1: Generate Clusters

```

PARTITION-CLUSTER( $C_1, C_2$ )
P1. repeat
P2.   for all  $x \in C_1 \cup C_2$  do
P3.     if cluster( $x$ ) =  $C_1$  then
P4.        $C_u \leftarrow C_1$ ;  $C_v \leftarrow C_2$ ;
P5.     else
P6.        $C_u \leftarrow C_2$ ;  $C_v \leftarrow C_1$ ;
P7.     end if
P8.      $Q_i \leftarrow Quality(C_u) + Quality(C_v)$ ;
P9.      $Q_s \leftarrow Quality(C_u - \{x\}) + Quality(C_v \cup \{x\})$ ;
P10.    if  $Q_s > Q_i$  then
P11.       $C_u.Remove(x)$ ;
P12.       $C_v.Insert(x)$ ;
P13.    end if
P14.  end for
P15. until  $C_1$  and  $C_2$  are stable
  
```

Figure 2: Partition Cluster

PARTITION-CLUSTER

The PARTITION-CLUSTER algorithm is given in Fig.2. The algorithm continuously evaluates, for each element $x \in C_1 \cup C_2$, to check whether a reassignment increases the homogeneity of the two clusters.

Lines P8 and P9 compute the involvement of x to the local quality in two cases: either x remains in its original

cluster (C_u) or x is moved to the other cluster (C_v). If moving x gives an improvement in the local quality, then the swapping is done (lines P10–P13). Lines P2–P14 in the algorithm is nested into a main loop: elements are continuously checked for swapping until a convergence is met. The splitting process can be sensitive to the order upon which elements are considered: In the first stage, it could be not convenient to reassign the generic x_i from C_1 to C_2 , whereas a convenience in performing the swap can be found after the relocation of some other element x_j . The main loop partly smoothes this effect by repeatedly relocating objects until convergence is met. Better PARTITION-CLUSTER can be made strongly insensitive to the order with which cluster elements are considered. The basic idea is discussed next. The elements that mostly influence the locality effect are either outlier transactions (that is, those containing mainly items, whose frequency within the cluster is rather low) or common transactions (which, dually, contain very frequent items). In the first case, C_2 is unable to attract further transactions, whereas in the second case, C_2 is likely to attract most of the transactions (and, consequently, C_1 will contain outliers).

The key idea is to rank and sort the cluster elements before line P1, which is on the basis of their splitting effectiveness. To this purpose, each transaction x belonging to cluster C can be associated with a weight $w(x)$, which indicates its splitting effectiveness. x is eligible for splitting C if its items allow us to divide C into two homogeneous sub-clusters. In this respect, the Gini index is a natural way to quantify the splitting effectiveness $G(a)$ of the individual attribute value $a \in x$. Precisely, $G(a) = 1 - \Pr(a|C)^2 - (1 - \Pr(a|C))^2$, where $\Pr(a|C)$ denotes the probability of a within C . $G(a)$ is close to its maximum whenever a is present in about half of the transactions of C and reaches its minimum whenever a is unfrequent or common within C . The overall splitting effectiveness of x can be defined by averaging the splitting effectiveness of its constituting items $w(x) = \text{avg}_{a \in x} (G(a))$. Once ranked, the elements $x \in C$ can be considered in descending order of their splitting effectiveness at line P2. This guarantees that C_2 is initialized with elements, which do not represent outliers and still are likely to be removed from C_1 . This removes the dependency on the initial input order of the data. With decision tree learning, MPC exhibits a preference bias, which is encoded within the notion of homogeneity and can be viewed as the preference for compact clustering trees. Indeed, due to the splitting effectiveness heuristic, homogeneity is enforced by the effects of the Gini index. At each split, this tends to isolate clusters of transactions with mostly frequent attribute values, from which the compactness of the overall clustering tree follows.

b. STABILIZE-CLUSTERS

PARTITION-CLUSTER improves the local quality of a cluster. And STABILIZE-CLUSTERS try to increase partition quality. It is carried out by finding the most suitable clusters for each element among the ones which are there in the partition.

Fig. 3 shows the pseudo code of the procedure. The central part of the algorithm is a main loop which (lines S2–

S17) examines all the available elements. For each element x , a pivot cluster is identified, which is the cluster containing x . Then, the available clusters are continuously evaluated. The insertion of x in the current cluster is done (lines S5–S6), and the updated quality is compared with the original quality.

```

STABILIZE-CLUSTERS( $P$ )
S1.  repeat
S2.    for all  $x \in D$  do
S3.       $C_{pivot} \leftarrow \text{cluster}(x)$ ;  $Q \leftarrow \text{Quality}(P)$ ;
S4.      for all  $C \in P$  do
S5.         $C_{pivot}.\text{REMOVE}(x)$ ;
S6.         $C.\text{INSERT}(x)$ ;
S7.        if  $\text{Quality}(P) > Q$  then
S8.          if  $C_{pivot} = \emptyset$  then
S9.             $P.\text{REMOVE}(C_{pivot})$ ;
S10.         end if
S11.         $C_{pivot} \leftarrow C$ ;  $Q \leftarrow \text{Quality}(P)$ ;
S12.       else
S13.         $C_{pivot}.\text{INSERT}(x)$ ;
S14.         $C.\text{REMOVE}(x)$ ;
S15.       end if
S16.     end for
S17.  end repeat
S18. until  $P$  is stable
    
```

Figure 3: Stabilize Clusters

If an improvement is obtained, then the swap is accepted (line S11). The new pivot cluster is the one now containing x , and if the removal of x makes the old pivot cluster empty, then the old pivot cluster is removed from the partition P . If there is no improvement in quality, x is restored into its pivot cluster, and a new cluster is examined. The main loop is iterated until a stability condition for clusters is achieved.

c. Cluster and Partition Qualities

AT-DC gives two different quality measures, 1) local homogeneity within a cluster and 2) global homogeneity of the partition. As shown in Fig. 1, it is noticed that partition quality is used for checking whether the insertion of a new cluster is really suitable: it is for maintaining compactness. Cluster quality in procedure PARTITIONCLUSTER is done for good separation.

Cluster quality is known when there is a high degree of intracluster homogeneity and intercluster homogeneity. As given in [35], there is strong relation between intracluster homogeneity and the probability $\text{Pr}(a_i|C_k)$ that item a_i appears in a transaction containing in C_k . There is a strong relationship between intercluster separation and $\text{Pr}(x \in C_k, a_i \in x)$. Cluster homogeneity and separation is computed by relating it to the unity of items within the transactions that it contains. Cluster quality is equal to the combination of the above probability,

$\sum_{a \in M_C} \text{Pr}(a|C) \text{Pr}(C|a) \text{Pr}(a)$. The last term is used for weighting the importance of item a in the summation: Essentially, high values from low-frequency items are less

relevant than those from high-frequency values. By the Bayes theorem, the above formula is expressed as $\text{Pr}(C) \sum_{a \in M_C} \text{Pr}(a|C)^2$ [33]. Terms $\text{Pr}(a|C)^2$ (relative strength of a within C) and $\text{Pr}(C)$ (relative strength of C) work in contraposition. It is easy to compute the gain in strength for each item with respect to the whole data set, that is

$$\text{Quality}(C_k) = \text{Pr}(C_k) \sum_{a \in M_{C_k}} [\text{Pr}(a|C_k)^2 - \text{Pr}(a|D)^2] \quad \dots\dots (1)$$

Where,

- C_k – cluster
- $\text{Pr}(C_k)$ – relative strength of C_k
- $a \in M_{C_k}$ – an item
- $M = \{a_1, \dots, a_m\}$ is set of Boolean attributes
- $\text{Pr}(a|C_k)$ – relative strength of a within C_k
- $\text{Pr}(a|D)$ – relative strength of a within D
- $D = \{x_1, \dots, x_n\}$ is data set of tuples defined on M

$$\text{Quality}(C_k) = \frac{n}{N} \sum_{a \in M_{C_k}} \left(\frac{n_a}{n} \right)^2 - \left(\frac{N_a}{N} \right)^2 \quad \dots\dots (2)$$

where n_a and N_a represent the frequencies of a in C and D , respectively. The value of $\text{Quality}(C_k)$ is updated as soon as a new transaction is added to C .

IV. RESULTS AND ANALYSIS

Two real-life data sets were evaluated. A description of each data set employed for testing is provided next, together with an evaluation of the MPC performances.

UCI DATASETS [34]

Zoo: Zoo dataset contains 103 instances, each having 18 attributes (animal name, 15 Boolean attributes and 2 numerics). The "type" attribute appears to be the class attribute. In total there are 7 classes of animals, that is, class 1 has 41 set of animals, class 2 has 20 set of animals, class 3 has 5 set of animals, class 4 has 13 set of animals, class 5 has 4 set of animals, class 6 has 8 set of animals and class 7 has 10 set of animals. Here is a breakdown of which animals are in which type: (it is unusual that there are 2 instances of "frog" and one of "girl"!). There are no missing values in this dataset. Table 1 shows that in cluster 1, a class 2 is having high homogeneity and in cluster 2, classes 3, 5 and 7 are having high homogeneity.

Hepatitis: Hepatitis contains 155 instances, each having 20 attributes. It represents the observation of patients. Each instance is one patient's record according to 20 attributes (for example, age, steroid, antivirals, and spleen palpable). Some attributes contains missing values. A class as "DIE" or

“LIVE” is given to each instance. Out of 155 instances, 32 are “DIE” and 123 are “LIVE”. Table 2 shows that in cluster 1 and cluster 2 are having high homogeneity. In cluster 2 and 4 there are 2 (DIE) and 1 (LIVE) instances which are misclassified.

Table 1: Confusion matrix for zoo

Cluster No.	Classes						
	1	2	3	4	5	6	7
1	17	20	0	5	0	2	0
2	24	0	5	8	4	6	10

Table 2: Confusion matrix for Hepatitis

Cluster No.	Classes	
	DIE	LIVE
1	17	0
2	2	63
3	0	59
4	13	1

V. CONCLUDING REMARK

This innovative MPC algorithm is fully-automatic, parameter-free approach to cluster high-dimensional categorical data. The main advantage of our approach is its capability of avoiding explicit prejudices, expectations, and presumptions on the problem at hand, thus allowing the data itself to speak. This is useful with the problem at hand, where the data is described by several relevant attributes.

A limitation of our proposed approach is that the underlying notion of cluster quality is not meant for catching conceptual similarities, that is, when distinct values of an attribute are used for denoting the same concept. Probabilities are provided to evaluate cluster homogeneity only in terms of the frequency of items across the underlying transactions. Hence, the resulting notion of quality suffers from the typical limitations of the approaches, which use exact-match similarity measures to assess cluster homogeneity. To this purpose, conceptual cluster homogeneity for categorical data can be easily added to the framework of the MPC algorithm.

Another limitation of our approach is that it cannot deal with outliers. These are transactions whose structure strongly differs from that of the other transactions being characterized by low-frequency items. A cluster containing such transaction exhibits low quality. Worst, outliers could negatively affect the PARTITION-CLUSTER procedure by preventing the split to be accepted (because of an arbitrary assignment of such outliers, which would lower the quality of the partitions). Hence, a significant improvement of MPC can be obtained by defining an outlier detection procedure that is

capable of detecting and removing outlier transactions before partitioning the clusters. The research work can be extended further to improve the quality of clusters by removing outliers.

REFERENCES

- [1] J. Grabmeier and A. Rudolph, “Techniques of Cluster Algorithms in Data Mining,” *Data Mining and Knowledge Discovery*, vol. 6, no. 4, pp. 303-360, 2002.
- [2] A. Jain and R. Dubes, *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [3] R. Ng and J. Han, “CLARANS: A Method for Clustering Objects for Spatial Data Mining,” *IEEE Trans. Knowledge and Data Eng.*, vol. 14, no. 5, pp. 1003-1016, Sept./Oct. 2002.
- [4] Z. Huang, “Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values,” *Data Mining and Knowledge Discovery*, vol. 2, no. 3, pp. 283-304, 1998.
- [5] C. Gozzi, F. Giannotti, and G. Manco, “Clustering Transactional Data,” *Proc. Sixth European Conf. Principles and Practice of Knowledge Discovery in Databases (PKDD '02)*, pp. 175-187, 2002.
- [6] S. Deerwester et al., “Indexing by Latent Semantic Analysis,” *J. Am. Soc. Information Science*, vol. 41, no. 6, 1990.
- [7] L. Parsons, E. Haque, and H. Liu, “Subspace Clustering for High-Dimensional Data: A Review,” *SIGKDD Explorations*, vol. 6, no. 1, pp. 90-105, 2004.
- [8] G. Gan and J. Wu, “Subspace Clustering for High Dimensional Categorical Data,” *SIGKDD Explorations*, vol. 6, no. 2, pp. 87-94, 2004.
- [9] M. Zaki and M. Peters, “CLICK: Mining Subspace Clusters in categorical Data via k-Partite Maximal Cliques,” *Proc. 21st Int'l Conf. Data Eng. (ICDE '05)*, 2005.
- [10] Y. Yang, X. Guan, and J. You, “CLOPE: A Fast and Effective Clustering Algorithm for Transactional Data,” *Proc. Eighth ACM Conf. Knowledge Discovery and Data Mining (KDD '02)*, pp. 682-687, 2002.
- [11] E. Han, G. Karypis, V. Kumar, and B. Mobasher, “Clustering in a High Dimensional Space Using Hypergraph Models,” *Proc. ACM SIGMOD Workshops Research Issues on Data Mining and Knowledge Discovery (DMKD '97)*, 1997.
- [12] M. Ozdal and C. Aykanat, “Hypergraph Models and Algorithms for Data-Pattern-Based Clustering,” *Data Mining and Knowledge Discovery*, vol. 9, pp. 29-57, 2004.
- [13] K. Wang, C. Xu, and B. Liu, “Clustering Transactions Using Large Items,” *Proc. Eighth Int'l Conf. Information and Knowledge Management (CIKM '99)*, pp. 483-490, 1999.
- [14] D. Barbara, J. Couto, and Y. Li, “COOLCAT: An Entropy-Based Algorithm for Categorical Clustering,” *Proc. 11th ACM Conf. Information and Knowledge Management (CIKM '02)*, pp. 582-589, 2002.
- [15] P. Andritsos, P. Tsaparas, R. Miller, and K. Sevcik, “LIMBO: Scalable Clustering of Categorical Data,” *Proc. Ninth Int'l Conf. Extending Database Technology (EDBT '04)*, pp. 123-146, 2004.
- [16] M.O.T. Li and S. Ma, “Entropy-Based Criterion in Categorical Clustering,” *Proc. 21st Int'l Conf. Machine Learning (ICML '04)*, pp. 68-75, 2004.
- [17] I. Cadez, P. Smyth, and H. Mannila, “Probabilistic Modeling of Transaction Data with Applications to Profiling, Visualization, and Prediction,” *Proc. Seventh ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '01)*, pp. 37-46, 2001.
- [18] M. Carreira-Perpinan and S. Renals, “Practical Identifiability of Finite Mixture of Multivariate Distributions,” *Neural Computation*, vol. 12, no. 1, pp. 141-152, 2000.
- [19] G. McLachlan and D. Peel, *Finite Mixture Models*. John Wiley & Sons, 2000.
- [20] M. Meila and D. Heckerman, “An Experimental Comparison of Model-Based Clustering Methods,” *Machine Learning*, vol. 42, no. 1/2, pp. 9-29, 2001.
- [21] J.G.S. Zhong, “Generative Model-Based Document Clustering: A Comparative Study,” *Knowledge and Information Systems*, vol. 8, no. 3, pp. 374-384, 2005.
- [22] A. Gordon, *Classification*. Chapman and Hall/CRC Press, 1999.

- [23] C. Fraley and A. Raftery, "How Many Clusters? Which Clustering Method? The Answer via Model-Based Cluster Analysis," The Computer J., vol. 41, no. 8, 1998.
- [24] P. Smyth, "Model Selection for Probabilistic Clustering Using Cross-Validated Likelihood," Statistics and Computing, vol. 10, no. 1, pp. 63-72, 2000.
- [25] D. Pelleg and A. Moore, "X-Means: Extending K-Means with Efficient Estimation of the Number of Clusters," Proc. 17th Int'l Conf. Machine Learning (ICML '00), pp. 727-734, 2000.
- [26] M. Sultan et al., "Binary Tree-Structured Vector Quantization Approach to Clustering and Visualizing Microarray Data," Bioinformatics, vol. 18, 2002.
- [27] S. Guha, R. Rastogi, and K. Shim, "ROCK: A Robust Clustering Algorithm for Categorical Attributes," Information Systems, vol. 25, no. 5, pp. 345-366, 2001.
- [28] J. Basak and R. Krishnapuram, "Interpretable Hierarchical Clustering by Constructing an Unsupervised Decision Tree," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 1, Jan. 2005.
- [29] H. Blockeel, L.D. Raedt, and J. Ramon, "Top-Down Induction of Clustering Trees," Proc. 15th Int'l Conf. Machine Learning (ICML'98), pp. 55-63, 1998.
- [30] B. Liu, Y. Xia, and P. Yu, "Clustering through Decision Tree Construction," Proc. Ninth Int'l Conf. Information and Knowledge Management (CIKM '00), pp. 20-29, 2000.
- [31] Yi-Dong Shen, Zhi-Yong Shen and Shi-Ming Zhang, "Cluster Cores – based Clustering for High – Dimensional Data".
- [32] Alexander Hinneburg and Daniel A. Keim, Markus Wawryniuk, "HD-Eye-Visual of High-Dimensional Data: A Demonstration".
- [33] http://en.wikipedia.org/wiki/Bayes'_theorem
- [34] UCI Machine Learning Repository <http://www.ics.uci.edu/~mllearn/>
- [35] D. Fisher, "Knowledge Acquisition via Incremental Conceptual Clustering," Machine Learning, vol. 2, pp. 139-172, 1987.

AUTHORS PROFILE



Sunita M. Karad has received B.E. degree in Computer Engineering from Marathwada University, India in 1992, M.E. degree from Pune University in 2007. She is a registered Ph.D. student of Amravati University. She is currently working as Assistant Professor in Computer Engineering department in MIT, Pune. She has more than 10 years of teaching experience and successfully handles administrative work in MIT, Pune. Her research interest includes

Data mining, Business Intelligence & Aeronautical space research.



Dr. Vijay M. Wadhai received his B.E. from Nagpur University in 1986, M.E. from Gulbarga University in 1995 and Ph.D. degree from Amravati University in 2007. He has experience of 25 years which includes both academic (17 years) and research (8 years). He has been working as a Dean of Research, MITSOT, MAE, Pune (from 2009) and simultaneously handling the post of Director - Research and Development, Intelligent Radio Frequency (IRF) Group, Pune (from 2009). He is currently guiding 12 students for their PhD work in both Computers and Electronics & Telecommunication area. His research interest includes Data Mining, Natural Language processing, Cognitive Radio and Wireless Network, Spectrum Management, Wireless Sensor Network, VANET, Body Area Network, ASIC Design - VLSI. He is a member of ISTE, IETE, IEEE, IES and GISFI (Member Convergence Group), India.

Dr. Madan U. Kharat has received his B.E. from Amravati University, India in 1992, M.S. from Devi Ahilya University (Indore), India in 1995 and Ph.D. degree from Amravati University, India in 2006. He has experience of 18 years in academics. He has been working as a Principle of PLIT, Yelgaon, Budhana. His research interest includes Deductive Databases, Data Mining and Computer Networks.



Prasad S. bachelor's degree Amravati Computer Science Engineering, currently a lecturer research interest and Data Mining, databases and



Halgaonkar received his in Computer Science from University in 2006 and M.Tech in from Walchand College of Shivaji University in 2010. He is in MITCOE, Pune. His current includes Knowledge discovery deductive databases, Web Semi-Structured data.



Body Area Network.

Dipti D. Patil has received B.E. degree in Computer Engineering from Mumbai University in 2002 and M.E. degree in Computer Engineering from Mumbai University, India in 2008. She has worked as Head & Assistant Professor in Computer Engineering Department in Vidyavardhini's College of Engineering & Technology, Vasai. She is currently working as Assistant Professor in MITCOE, Pune. Her Research interests include Data mining, Business Intelligence and

A Pilot Based RLS Channel Estimation for LTE SC-FDMA in High Doppler Spread

M. M. Rana

Department of Electronics and Communication Engineering
Khulna University of Engineering and Technology
Khulna, Bangladesh

Abstract—Main challenges for a terminal implementation are efficient realization of the inner receiver, especially for channel estimation (CE) and equalization. In this paper, pilot based recursive least square (RLS) channel estimator technique is investigated for a long term evolution (LTE) single carrier-frequency division multiple access (SC-FDMA) system in high Doppler spread environment. This CE scheme uses adaptive RLS estimator which is able to update parameters of the estimator continuously, so that knowledge of channel and noise statistics are not required. Simulation results show that the RLS CE scheme with 500 Hz Doppler frequency has 3 dB better performances compared with 1.5 kHz Doppler frequency.

Keywords— Channel estimation, LTE, RLS, SC-FDMA.

I. INTRODUCTION

The 3rd generation partnership project (3GPP) members started a feasibility study on the enhancement of the universal terrestrial radio access (UTRA), to improve the mobile phone standard to cope with future requirements. This project was called long term evolution (LTE) [1], [2]. LTE uses orthogonal frequency division multiple access (OFDMA) for downlink and single carrier-frequency division multiple access (SC-FDMA) for uplink transmission [1]. A highly efficient way to cope with the frequency selectivity of wideband channel is OFDMA. OFDMA is an effective technique for combating multipath fading and for high bit rate transmission over mobile wireless channels. Channel estimation (CE) has been successfully used to improve the system performance. It can be employed for the purpose of detecting received signal, improve signal-to-noise ratio (SNR), channel equalization, cochannel interference (CCI) rejection, and improved the system performance [3-5].

In general, CE techniques can be divided into three categories such as pilot CE, blind CE, and semi-blind CE [10], [11]. Pilot CE techniques offer low computational complexity and good performance [12]. The blind CE techniques exploit the statistical behavior of the received signals and require a large amount of data [13]. Semi-blind CE methods are used a combination of data aided and blind methods [11]. The pilot CE algorithm requires probe sequences; the receiver can use this probe sequence to reconstruct the transmitted waveform

[6-8]. Pilot symbols can be placed either at the beginning of each burst as a preamble or regularly through the burst. Pilot sequences are transmitted at certain positions of the SC-FDMA frequency time pattern, in its place of data.

Adaptive CE has been, and still is, an area of active research topics, playing imperative roles in an ever growing number of applications such as wireless communications where the channel is rapidly time-varying. Signal processing techniques that use recursively estimated, time varying models are normally called adaptive. Different adaptive CE algorithms have been proposed over the years for the purpose of updating the channel coefficient. The least mean square (LMS) method, its normalized version (NLMS), the affine projection algorithm (APA), as well as the recursive least square (RLS) method are well known examples of such CE algorithms. The well known LMS/NLMS CE algorithms are attractive from a computational complexity point of view but their convergence behavior for highly correlated input signals is poor. The RLS CE method resolves this trouble, but at the expense of increased complexity. A very large number of fast RLS CE methods have been developed over the years, but regrettably, it seems that the better a fast RLS CE method is in terms of computational efficiency and numerical stability. In addition, the RLS algorithm has the recursive inversion of an estimate of the autocorrelation matrix of the input signal as its cornerstone, problems arise, if the autocorrelation matrix is rank deficient.

In this paper, we investigate the adaptive RLS CE method in the LTE SC-FDMA systems in high Doppler spread environment. This CE method uses adaptive estimator which is able to update parameters of the estimator continuously so that knowledge of channel and noise statistics are not required. Simulation results show that the RLS CE scheme with 500 Hz Doppler frequency has 3 dB better performances compared with 1500 Hz Doppler frequency.

We use the following notations throughout this paper: bold face lower letter is used to represent vector. Superscripts \mathbf{x}^* and \mathbf{x}^T denote the conjugate and conjugate transpose of the complex vector \mathbf{x} respectively.

The remainder of the paper is organized as follows: section II describes wireless communication systems and LTE SC-FDMA systems model is describes in section III. The RLS CE scheme is presented in section IV, and its performance is

analyzed in section V. Finally, some concluding remarks are given in section VI.

II. WIRELESS COMMUNICATION SYSTEMS

Nowadays, cellular mobile phones have become an important tool and part of daily life. In the last decade, cellular systems have experienced fast development and there are currently about two billion users over the world. The idea of cellular mobile communications is to divide large zones into small cells, and it can provide radio coverage over a wider area than the area of one cell. This concept was developed by researchers at AT & T Bell laboratories during the 1950s and 1960s. The initial cellular system was created by nippon telephone & telegraph (NTT) in Japan, 1979. From then on, the cellular mobile communication has evolved.

The mobile communication systems are frequently classified as different generations depending of the service offered. The first generation (1G) comprises the analog communication techniques, and it was mainly built on frequency modulation (FM) and frequency division multiple accesses (FDMA). Digital communication techniques appeared in the second generation (2G) systems, and main access schemes are time division multiple access (TDMA) and code division multiple access (CDMA). The two most commonly accepted 2G systems are global system for mobile (GSM) and interim standard-95 (IS-95). These systems mostly offer speech communication, but also data communication limited to rather low transmission rates. The concept of the third generation (3G) system started operations on October, 2002 in Japan. The 3GPP members started a feasibility study on the enhancement of the universal terrestrial radio access (UTRA) in December 2004, to improve the mobile phone standard to cope with future requirements. This project was called LTE. LTE uses SC-FDMA for uplink transmission and OFDMA for downlink transmission. Fig. 1 summarizes the evolution path of cellular mobile communications systems.

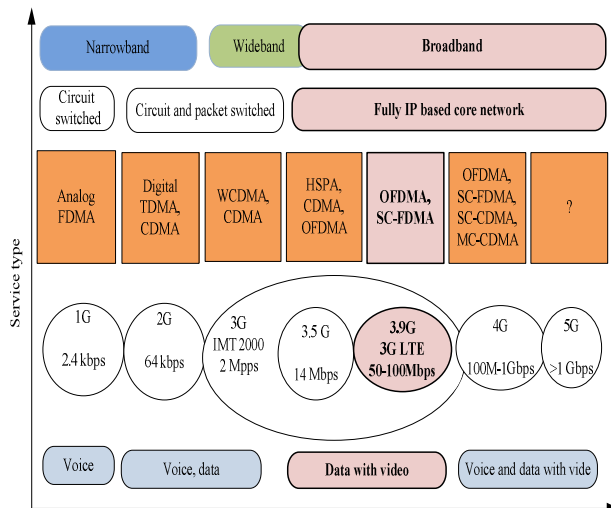


Fig. 1. Evolution path in mobile communication systems.

III. LTE SC-FDMA SYSTEMS DESCRIPTION

In this section, we briefly explain LTE SC-FDMA system model, fading channel statistics, and received signal model.

A. Baseband system model

A baseband block diagram for the communications system under investigation is shown in Fig. 2.

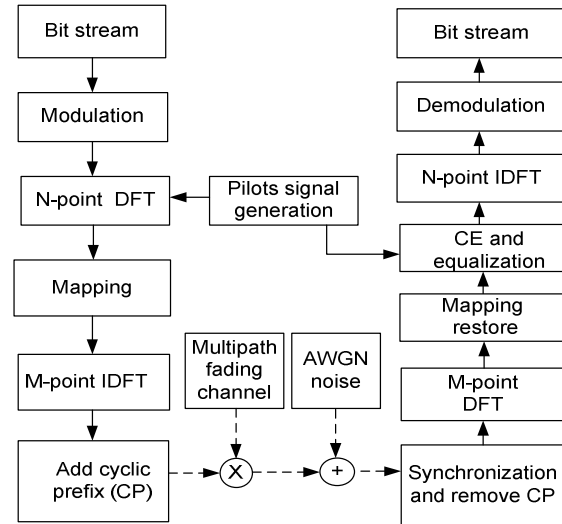


Fig. 2. Block diagram of a LTE SC-FDMA system.

At the transmitter, a baseband multiple phase shift keying modulator takes binary sequence and produces the signaling waveforms

$$\begin{aligned} m_i(t) &= \sqrt{\frac{2E}{T}} \cos(\omega t + \alpha_i), 0 < t < T \\ &= \sqrt{\frac{2E}{T}} [\cos(\alpha_i) \cos(\omega t) - \sin(\alpha_i) \sin(\omega t)] \\ &= a_i b(t) + c_i d(t), \end{aligned} \quad (1)$$

where T is the symbol duration, E is the energy of $m_i(t)$,

$\omega = 2\pi f$, f is the carrier frequency, phase angle

$$\alpha = \frac{2\pi i}{M}, M \text{ is the alphabate size, } \alpha_i = \sqrt{E} \cos \alpha_i$$

inphase basis, $b(t) = \sqrt{\frac{2}{T}} \cos(\omega t)$, $c_i = \sqrt{E} \sin \alpha_i$, and

quadrature basis, $d(t) = -\sqrt{\frac{2}{T}} \sin(\omega t)$. CE is often achieved

by multiplexing known symbols, so called, pilot symbols into data sequences [1]. These modulated symbols and pilots perform N-point discrete Fourier transform (DFT) to produce a frequency domain representation:

$$s_i(t) = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} m_i(t) e^{-\frac{j2\pi mt}{N}}, \quad (2)$$

where j is the imaginary unit. It then maps each of the N -point DFT outputs to one of the orthogonal sub-carriers mapping that can be transmitted. There are two principal sub-carrier mapping modes: localized mode, and distribution mode. In distributed sub-carrier mode, the outputs are allocated equally spaced sub-carrier, with zeros occupying the unused sub-carrier in between. While in localized sub-carrier mode, the outputs are confined to a continuous spectrum of sub-carrier. Interleaved sub-carrier mapping mode of FDMA (IFDMA) is another special sub-carrier mapping mode [13], [14]. The difference between DFDMA and IFDMA is that the outputs of IFDMA are allocated over the entire bandwidth, whereas the DFDMA outputs are allocated every several subcarriers [15], [16].

Finally, the inverse DFT (IDFT) module output is followed by a cyclic prefix (CP) insertion that completes the digital stage of the signal flow. The CP is used to eliminate ISI and preserve the orthogonality of the tones. Assume that the channel length of CP is larger than the channel delay spread [17].

B. Channel model

Channel model is a mathematical representation of the transfer characteristics of the physical medium. These models are formulated by observing the characteristics of the received signal. According to the documents from 3GPP, a radio wave propagation can be described by multipaths which arise from reflection and scattering [17]. The received signal at the mobile terminal is a superposition of all paths. If there are L distinct paths from transmitter to the receiver, the impulse response of the multipath fading channel can be represented as [17]:

$$\alpha(m, \tau) = \sum_{j=1}^L \omega_j(m) \delta[m - \tau_j(m)], \quad (3)$$

where $\omega_j(m)$ and $\tau_j(m)$ are attenuations and delays for each path at time instant m , and $\delta(\cdot)$ is the Dirac delta function. The fading process for the mobile radio channel is given by

$$\omega(v) = \omega_j \sqrt{1 - (v/f_d)}, \quad (4)$$

where Doppler frequency $f_d = s/\lambda$, s is the speed of the mobile, and λ is the wavelength of the transmitted carrier. In order to do simulations as close to the reality as possible, it is essential to have a good channel model. This model is used to describe the fast variations of the received signal strength due to changes in phases when a mobile terminal moves. In case of wideband modeling, each path of the impulse response can be modeled as Rayleigh distributed with uniform phase except line of sight (LOS) component cases [17].

C. Received signal model

The transmitted symbols propagating through the radio channel can be modeled as a circular convolution between the CIR and the transmitted data block i.e., $[s(m) * \omega(m, \tau)]$.

Since, the channel coefficient is usually unknown to the receiver, it needs to be efficiently estimated. The impulse response of multipath fading channel can be represented by a tap-delayed line filter with time varying coefficients and symbol-rate spaced coefficients.

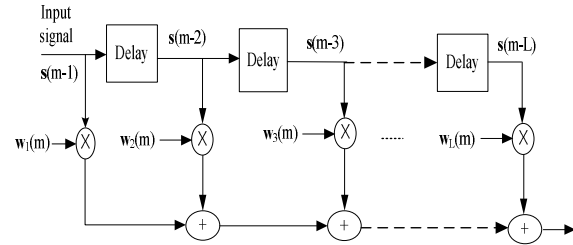


Fig. 3. L-tapped delay line filter of a fading channel.

At the receiver, the opposite set of the operation is performed. After synchronization, CP samples are discarded and the remaining samples are processed by the DFT to retrieve the complex constellation symbols transmitted over the orthogonal sub-channels. The received signals are demapped and equalizer is used to compensate for the radio channel frequency selectivity. After IDFT operation, these received signals are demodulated and soft or hard values of the corresponding bits are passed to the decoder. The decoder analyzes the structure of received bit pattern and tries to reconstruct the original signal.

IV. RLS ADAPTIVE CE METHOD

An adaptive CE technique is a process that changes its parameters as it gain more information of its possibly changing environment. Among many iterative techniques that exist in the open literature, the well-liked classes of approaches which are achieve from the minimization of the mean square error (MSE) between the output of the adaptive filter and desired signal to perform CE as shown in Fig. 4.

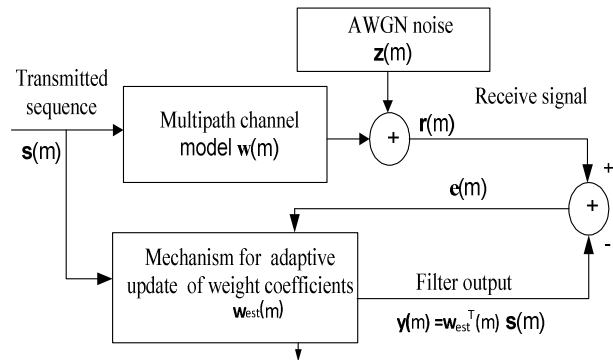


Fig. 4. Scheme for adaptive CE.

The signal $s(m)$ is transmitted via a time-varying channel $w(m)$, and corrupted by an additive noise estimated by using any kind of CE method. The main aim of most channel estimation algorithms is to minimize the MSE i.e., between the received signal and its estimate, while utilizing as little

computational resources as possible in the estimation process. In the Fig. 4, we have unknown multipath fading channel, that has to be estimated with an adaptive filter whose weight are updated based on some criterion so that coefficients of adaptive filter should be as close as possible to the unknown channel. The RLS CE requires all the past samples of the input and the desired output is available at each iteration. The objective function of a RLS CE algorithm is defined as an exponential weighted sum of errors squares:

$$c(m) = \sum_{n=1}^m \lambda^{n-m} \mathbf{e}^H(m) \mathbf{e}(m) + \delta \lambda^n \mathbf{w}^H(m) \mathbf{w}(m), \quad (5)$$

where δ is a positive real number called regularization parameter, $\mathbf{e}(m)$ is the prior estimation error, and λ is the exponential forgetting factor with $0 < \lambda < 1$. The λ is used to ensure that data in the distant past are paid less concentration in order to provide the filter with estimating facility when it operates in a time varying environment. When $\lambda = 1$, the algorithm has growing memory because the values of the filter coefficients are a function of all the precedent input. In this case, all the values of the error signal, from the time the filter starts its process to the present, have the same influence on the cost function. Consequently, the adaptive filter losses its estimating ability, which is not important if the filter is used in a stationary environment. In contrast, when $0 < \lambda < 1$, the algorithm has exponentially decaying memory as the recent values of the observations have greater influence on the formation of the filter coefficients and tends to forget the old ones as shown in Fig. 5.

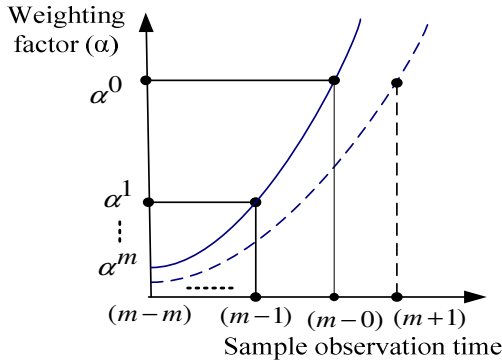


Fig. 5. Exponential weighting of observations at different time index.

The prior estimation error is the difference between the desired response and estimation signal:

$$\mathbf{e}(m) = \mathbf{h}(m) - \mathbf{w}^H(m) \mathbf{s}(m) \quad (7)$$

The objective function is minimized by taking the partial derivatives with respect to $\mathbf{w}(n)$ and setting the results equal to zero.

$$\begin{aligned} \frac{\delta c(m)}{\delta \mathbf{w}(m)} &= 0 = -2 \sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) \mathbf{e}^H(m) + 2\delta \lambda^n \mathbf{w}(m) \\ &= -2 \sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) [\mathbf{h}(m) - \mathbf{w}^H(m) \mathbf{s}(m)]^H + 2\delta \lambda^n \mathbf{w}(m) \\ \mathbf{w}(m) \left[\sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) \mathbf{s}^H(m) + \delta \lambda^n \mathbf{I} \right] &= \sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) \mathbf{h}^H(m) \\ \mathbf{R}_s(m) \mathbf{w}(m) &= \mathbf{R}_{sh}(m) \\ \mathbf{w}(m) &= \mathbf{R}_s^{-1}(m) \mathbf{R}_{sh}(m) \end{aligned} \quad (8)$$

where $\mathbf{R}_s(m)$ is the transmitted auto-correlation matrix

$$\mathbf{R}_s(m) = \sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) \mathbf{s}^H(m) + \delta \lambda^n \mathbf{I} = \lambda \mathbf{R}_s(m-1) + \mathbf{s}(m) \mathbf{s}^H(m)$$

and $\mathbf{R}_{sh}(m)$ is the cross correlation matrix i.e.,

$$\mathbf{R}_{sh}(m) = \sum_{n=1}^m \lambda^{n-m} \mathbf{s}(m) \mathbf{h}^H(m) = \lambda \mathbf{R}_{sh}(m-1) + \mathbf{s}(m) \mathbf{h}^H(m).$$

According to the Woodbury identity, the above $\mathbf{R}_{sh}(m)$ can be written as

$$\mathbf{R}_{sh}^{-1}(m) = \lambda^{-1} \mathbf{R}_{sh}^{-1}(m-1) - \frac{\lambda^{-2} \mathbf{R}_{sh}^{-1}(m-1) \mathbf{s}(m) \mathbf{s}^H(m) \mathbf{R}_{sh}^{-1}(m-1)}{1 + \lambda^{-1} \mathbf{s}^H(m) \mathbf{R}_{sh}^{-1}(m-1) \mathbf{s}(m)} \quad (9)$$

For convenience of computing, let $\mathbf{D}(m) = \mathbf{R}_{sh}(m)$ and

$$\mathbf{K}(m) = \frac{\lambda^{-1} \mathbf{D}(m-1) \mathbf{s}(m)}{1 + \lambda^{-1} \mathbf{s}^H(m) \mathbf{D}(m-1) \mathbf{s}(m)} \quad (10)$$

The $\mathbf{K}(m)$ is referred as a gain matrix. We may rewrite (9) as:

$$\mathbf{D}(m) = \lambda^{-1} \mathbf{D}(m-1) - \lambda^{-1} \mathbf{K}(m) \mathbf{s}^H(m) \mathbf{D}(m-1) \quad (11)$$

So simply (10) to

$$\mathbf{K}(m) = \mathbf{D}(m) \mathbf{s}(m) = \mathbf{R}_{sh}^{-1}(m) \mathbf{s}(m) \quad (12)$$

Substituting (11), (12) into (8), we obtain the following RLS CE formula

$$\begin{aligned} \mathbf{w}(m) &= \mathbf{w}(m-1) + \mathbf{K}(m) [\mathbf{h}(m) - \mathbf{w}^H(m-1) \mathbf{s}(m)]^H \\ &= \mathbf{w}(m-1) + \mathbf{K}(m) \mathbf{e}^H(m), \end{aligned} \quad (13)$$

where $\mathbf{e}(m)$ is a prior estimation error as

$$\mathbf{e}(m) = \mathbf{h}(m) - \mathbf{w}^H(m-1) \mathbf{s}(m) \quad (14)$$

Therefore equation (13) is the recursive RLS CE algorithm to update channel coefficient.

V. PERFORMANCE ANALYSIS

A. Complexity Analysis

The complexity of CE algorithm is of vital importance especially for time varying wireless communication channels, where it has to be performed periodically or even continuously. Table I summarizes the computational complexity of the RLS CE technique, where L is the channel length, and real number indicates scalar operation. Here we assume that each iteration requires the evaluation of the inner product $\mathbf{D}(m)\mathbf{h}(m)$ between two vectors of size L each. This calculation requires L multiplications and $L-1$ additions. Also assumed that the evaluation of the scalar addition or subtraction needs one real addition and multiplying the scalar by the vector requires L multiplications.

TABLE I
COMPLEXITY PER ITERATION

Operation	Complexity
Division	1
Multiplication	$L^2 + 5L + 1$
Addition	$L^2 + 3L$

B. Experimental results

The error performance of the aforementioned iterative estimation algorithm is explored by performing extensive computer simulations. All simulation parameters of the LTE SC-FDMA system in Doppler spread environments are summarized in Table II.

Table II
THE SYSTEMS PARAMETERS FOR SIMULATION

Systems parameters	Assumptions
System bandwidth	5 MHz
Sampling frequency	7.68 MHz
Sub-carrier spacing	9.765 kHz
Modulation data type	BPSK
FFT size	16
Sub-carrier mapping scheme	IFDMA
IFFT size	512
Data block size	32
Cyclic prefix	4 μ s
Channel	Rayleigh fading
Forgetting factor	0.99
Equalization	ZF
Doppler frequency	100, and 1.5 kHz

In practice, the perfect channel coefficient is unavailable, so estimated channel coefficient must be used instead. The more correct estimated channel coefficient is, the better MSE performance of the CE will achieve. Fig. 6 shows the MSE

versus SNR for the RLS CE method with different Doppler frequencies 500Hz and 1.5kHz. One can observed that the RLS CE method with 500 Hz Doppler frequency has 3 dB better performances compared with 1.5kHz Doppler frequency as desired. This CE scheme uses adaptive RLS estimator which is able to update parameters of the estimator continuously, so that knowledge of channel and noise statistics are not required. The similar behavior can be observed for BER performance in Fig. 7.

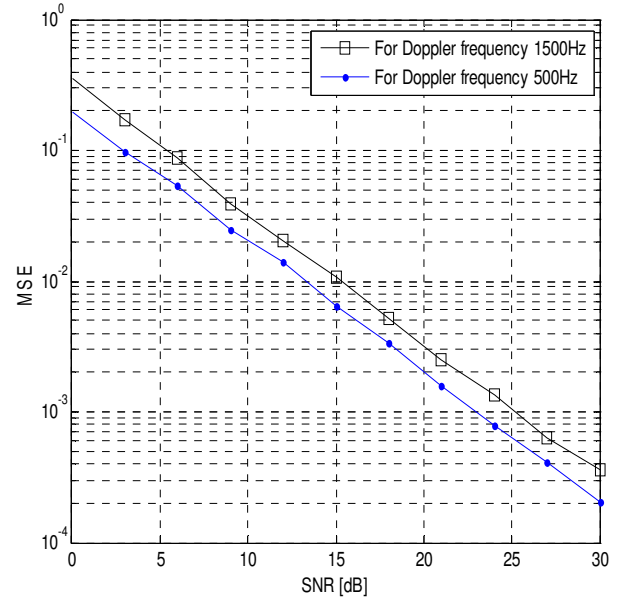


Fig. 6. MSE performance comparisons of the LMS CE method.

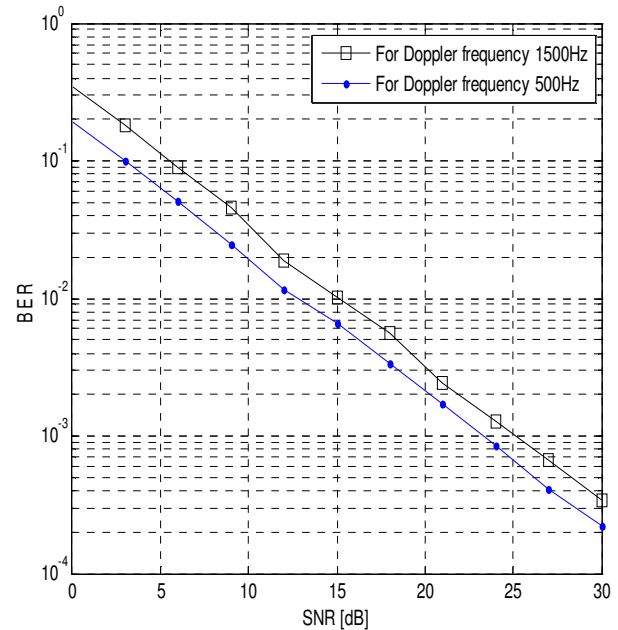


Fig.7. BER performance comparisons of the LMS CE method.

VI. CONCLUSION

In this paper, we explore the performance of RLS CE method for LTE SC-FDAM wireless communication systems with different Doppler frequencies. The complexities, MSE and BER performance of the RLS CE method, are analyzed and compared with the different Doppler frequencies. We can come to the conclusion that the RLS CE method with 500 Hz Doppler frequency has 3 dB superior performances compared with 1.5 kHz Doppler frequency.

REFERENCES

- [1] B. Karakaya, H. Arslan, and H. A. Cirpan, "Channel estimation for LTE uplink in high Doppler spread," *Proc. Int. Con. on WCNC*, pp. 1126-1130, April 2008.
- [2] J. Berkmann, C. Carbonelli, F. Dietrich, C. Drewes, and W. Xu, "On 3G LTE terminal implementation standard, algorithms, complexities and challenges," *Proc. Int. Con. on WCMC*, pp. 970-975, August 2008.
- [3] A. Ancora, C. Bona, and D.T.M. Slock, "Down-sampled impulse response least-squares channel estimation for LTE OFDMA," *Proc. Int. Con. ASSP*, Vol. 3, pp. 293-296, April 2007.
- [4] L. A. M. R. D. Temino, C. N. I. Manchon, C. Rom, T. B. Sorensen, and P. Mogensen, "Iterative channel estimation with robust wiener filtering in LTE downlink," *Proc. Int. Con. on VTC*, pp. 1-5, September 2008.
- [5] S. Y. Park, Y. Gu. Kim, and C. Gu. Kang, "Iterative receiver for joint detection and channel estimation in OFDM systems under mobile radio channels," *IEEE Trans. On Comm.*, vol. 53, iIssue 2, pp. 450-460, March 2004.
- [6] S. Haykin, "Adaptive Filter Theory," *Prentice-Hall International Inc.*, 1996.
- [7] J. J. V. D. Beek, O. E. M. Sandell, S. K. Wilson, and P. O. Boorjesson, "On channel estimation in OFDM systems," *Proc. Int. Con. on VTC*, vol. 2, pp. 815-819, July 1995.
- [8] O. Edfors, M. Sandell, J. V. D. Beek, and S. Wilson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. on Comm.*, vol. 46, no. 7, pp. 931-939, July 1998.
- [9] M.H. Hsieh, and C.H. Wei, "Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels," *IEEE Trans. on Consumer Electronics*, vol. 44, issue 1, pp. 217-225, February 1998.
- [10] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot symbol aided channel estimation by wiener filtering," *Proc. Int. Con. on ASSP*, pp. 1845-1848, vol.3, April 1997.
- [11] M. M. Rana, J. Kim, and W. K. Cho, "Low complexity downlink channel estimation for LTE systems," in *Proc. Int. Con. On Advanced Commun. Technology*, February 2010, pp. 1198-1202.
- [12] M. M. Rana, J. Kim, and W. K. Cho, "Performance Analysis of Sub-carrier Mapping in LTE Uplink Systems," in *Proc. Int. Con. On COIN*, August 2010.
- [13] F. Adachi, H. Tomeba, and K. Takeda, "Frequency-domain equalization for broadband single-carrier multiple access," *IEICE Trans. on Commun.*, vol. E92-B, no. 5, pp. 1441-1456, May 2009.
- [14] S. Yameogo, J. Palicot, and L. Cariou, "Blind time domain equalization of SC-FDMA signal," in *Proc. Vehicular Technology Conference*, pp. 1-4, September 2009.
- [15] S. H. Han and J. H. Lee, "An overview of peak to average power ratio reduction techniques for multicarrier transmission," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 2, pp. 56-65, 2005.
- [16] H. G. Myung, J. Lim, and D. J. Goodman, "Peak-to-average power ratio of single carrier FDMA signals with pulse shaping," in *Proc. Personal, Indoor and Mobile Radio Commun.*, September 2006.
- [17] W. C. Jakes, Ed., *Microwave mobile communications*. New York: Wiley-IEEE Press, Jan. 1994.

Priority Based Congestion Control for Multimedia Traffic In 3G Networks

Neetu Sharma¹, Amit Sharma², V.S Rathore³, Durgesh Kumar Mishra⁴

¹²³Department of Computer Engineering, Rajasthan, India

¹³Rajasthan College of Engineering for women, Rajasthan, India

²Shri Balagi College of Engineering & Technology, Rajasthan, India

⁴Acropolis Institute of Technology and Research, Indore, MP, India

ABSTRACT- There is a growing demand for efficient multimedia streaming applications over the Internet and next generation mobile networks. Multimedia streaming services are receiving considerable interest in the mobile network business. As communication technology is being developed, the user demand for multimedia services raises. The third generation (3G) mobile systems are designed to further enhance the communication by providing high data rates of the order of 2 Mbps. High Speed Downlink Packet Access (HSDPA) is an enhancement to 3G networks that supports data rates of several Mbit/s, making it suitable for applications like multimedia, in addition to traditional services like voice call. Services like person-to-person two way video calls or one way video calls, aim to improve person-to-person communication. Entertainment services like gaming, video streaming of a movie, movie trailers or video clips are also supported in 3G. Many more of such services are possible due to the augmented data rates supported by the 3G networks and because of the support for Quality of Service (QoS) differentiation in order to efficiently deliver required quality for different types of services.

This paper present congestion control schemes that are suitable for multimedia flows. The problem is that packet losses, during bad radio conditions in 3G, not only degrade the multimedia quality, but render the current congestion control algorithms as inefficient. This paper proposed a solution that integrated the congestion control schemes with a priority based multimedia packets to increase the speed of multimedia data and reduce the packet loss that is developed due to congestion in networks

Key words: UMTS, CN, BS, TFMCC, UTRAN, RNC

I. INTRODUCTION

The emerging multimedia application requires a fresh approach for congestion control. A widely popular congestion control schemes are

TCP friendly rate control (TFRC) and Adaptive increase multiplicative decrease (AIMD) used in networks. These algorithms used for multimedia traffic but not much effective in packet loss. TCP is the dominant transport protocol in the Internet, and the current stability of the Internet depends on its end-to-end congestion control, which uses an Additive Increase Multiplicative Decrease (AIMD) algorithm. End-to-end congestion control of best-effort traffic is required to avoid the congestion collapse of the global Internet [11]. While TCP congestion control is appropriate for applications such as bulk data transfer, some real-time applications (that is, where the data is being played out in real-time) find halving the sending rate in response to a single congestion indication to be unnecessarily severe. For providing a better congestion control with higher data rates a new effective scheme is used. Congestion control is an important issue in both wired and wireless streaming applications. Multimedia applications should use some form of congestion control, both in wired and cellular networks, in order to adapt the sending rate to the available bandwidth. Today's Internet stability is due to TCP and its congestion control algorithm. TCP represents a very efficient transport protocol in general and is suitable for data transfer. However, it has been argued [13] that TCP is unsuitable for video streaming because strict delay and jitter requirements of video streaming are not respected by TCP. Moreover, some TCP retransmissions are unnecessary for video when data may miss the arrival deadline and become obsolete. This has led researchers to look for alternative options. Most of the work related to congestion control for video flows has either emulated TCP or has used the TCP model. The well-known TCP-Friendly Rate Control (TFRC) congestion control consists in an equation based rate control mechanism [13][14][15], designed to keep a relatively steady sending rate while still being responsive to congestion. When used over wireless links, TFRC and TCP cannot distinguish between the wireless losses and the congestion losses. They both may suffer from the link underutilization if the connection traverses a wireless link.

This is because they consider dropped packets as a sure sign of congestion and reduce the ending rate significantly. The inability to identify a wireless loss followed by unnecessary reduction in sending rate results in link underutilization.

A. UMTS Introduction

Universal Mobile Telecommunications System (UMTS) is a third-generation (3G), wireless cellular network that uses Wideband Code Division Multiple Access (WCDMA) as its radio interface technology. UMTS offers higher data rates with respect to older 2G and 2.5G networks and, with the Release 5 version, is evolving into an all-IP, wireless network. The increased bandwidth provided by UMTS allows for the deployment of a wide range of services, like voice, data and multimedia streaming services. In wireless networks, congestion control, alone, may not be enough to ensure good quality of multimedia streaming and efficient utilization of the network. Packet losses due to the high bit error rate not only degrade the multimedia quality, but render the current congestion control algorithms as inefficient: these algorithms back-off on every packet loss even when there is no congestion. We integrate the congestion control schemes with an adaptive retransmission scheme in order to selectively retransmit some lost multimedia packets. Fig.1 shows the transmission of multimedia data over a wireless channel.

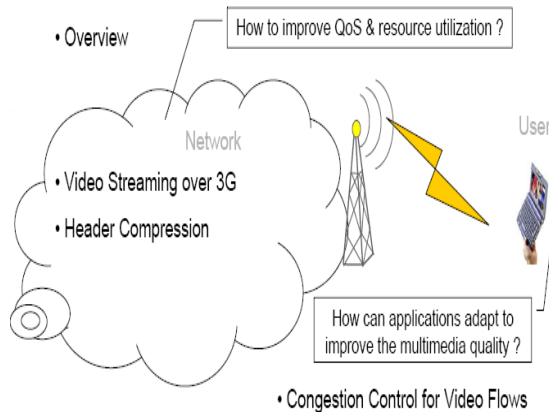


Fig. 1 Transmission of Multimedia data

B. GENERAL UMTS NETWORK:

UMTS, the successor of GSM, is evolving toward a future wireless all-IP network. In this paper we present how it supports real-time IP multimedia services, as these services are expected to drive the adoption of wireless all-IP networks.

UMTS networking architecture is organized in two domains. The user equipment (UE) and the public land mobile network (PLMN). The UE is used by the subscriber to access the UMTS services. PLAN is further divided into two land-based infrastructures

- (i) UTRAN (UMTS terrestrial radio- access network)
- (ii) CN (core network).

The UTRAN handles all radio-related functionalities and the CN is responsible for maintaining subscribers data and for connections. UTRAN contains two types of nodes: Radio network controller (RNC) and Node B. Node B is the base station and provides radio coverage to one or more cells. Node B is connected to UE via Uu interface and to the RNC via Iub interface. Uu is a radio interface based on the wideband code division multiple access (WCDMA) technology [7].

The CN consists of two types of general packet radio service support nodes (GSNs). That is gateway GSN (GGSN) and serving GSN (SGSN). SGSN provides the routing functionality. It manages a group of RNCs and interacts with the home location register (HLR). HLR permanently stores the subscriber data. SGSN is connected to GGSN via the Gn interface. RNC connects to SGSN via Iu interface. Through the GGSN, the UMTS network connects to external packet data networks like the internet.

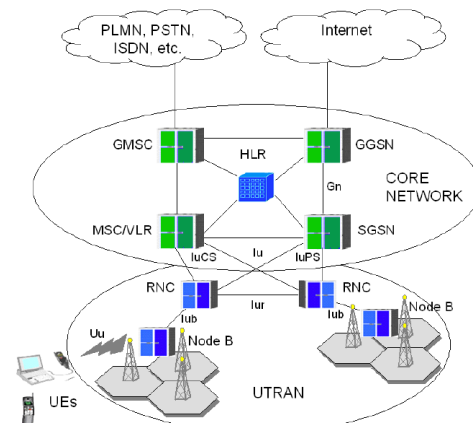


Fig.2 General UMTS Network

C. 3G/UMTS Problems

- Problems due to the use of IP
 - IP doesn't support real time streaming requirements
 - Overhead due to packet header
- Problems due to radio conditions
 - Scarce and time varying bandwidth
 - Congestion, wireless losses & large delay

D. UMTS QoS Classes

UMTS defines four QoS classes [2] and the classified traffic gets the treatment inside the UMTS network according to its class. The four QoS classes are:

- **Conversational class:** The traffic from the applications like person-to-person video or voice call is classified into conversational class. The delay and jitter requirements for this type of traffic are very strict. This is because on the both end points there is a human expecting the delivery of the voice and/or video data in very short time after it is sent.
- **Streaming class:** Video on Demand (VoD) falls under this class. The delay requirements are there but are not as strict as the conversational class.
- **Interactive class:** The interactive traffic like interactive e-mail or web browsing falls under this category. Though there is still some delay requirement, it is less strict than the conversational and streaming classes. Moreover, since the traffic mostly pertains to data applications, the bit error rate should be very low.
- **Background class:** This class is the most insensitive to delay. It includes the traffic from background applications like background email and SMS. Though, the bit error rate, like the Interactive class, should be very low.

D Congestion Control for Multimedia data

TCP-Friendly Rate Control (TFRC) is an end-to-end congestion control mechanism, whose goal is to provide rate control for unicast flows in IP networks. The main feature of TFRC is its ability to smoothly adapt the sending rate of a flow to network conditions, while competing for bandwidth with TCP flows in a relatively fair manner. TFRC was designed to offer a more stable sending rate than TCP on wired, best-effort networks, making it suitable for applications like multimedia streaming. We evaluate the performance of TFRC, compare it with that of TCP and new TFRC for different multimedia classes, under different scenarios of MAC-layer scheduling, radio conditions and background traffic.

TFRC [4][10] is an end-to-end congestion control mechanism suitable for applications with constraints on rate stability, like voice or streaming media. It has been designed to adapt the sending rate of a flow in a smooth manner, while trying to fairly share the available bandwidth with competing TCP flows. TFRC is an Internet standard [4], and it has been adopted at the IETF as one of the congestion

control profiles that may be used with the DCCP transport protocol [10]; TFRC may also be implemented by UDP-based applications wishing to perform congestion control. This paper presents a simulation study of TFRC over UMTS networks supporting

HSDPA. Since we are interested in video streaming applications, we evaluate the performance of TFRC in terms of rate stability over different time scales, and compare it with that of TCP. Several scenarios of MAC-layer scheduling, radio conditions and background traffic are considered.

This paper proposed a more reliable algorithm that provides congestion control for different multimedia classes. Priority assigned to each of the packet according to multimedia classes. So whenever the congestion occurs in the network the lowest priority packets are dropped. If overall loss rate for lower priority packets is not very high, then we can safely assume that the congestion loss rate for the highest priority packets will be insignificant. In such a case, the loss of highest priority packets will be mainly due to wireless errors. Thus, it is to be expected that, in general, there is a good correlation between wireless packet loss rate and the total loss rate of highest priority packets.

2. THE PROPOSED SCHEME

This paper provides a mechanism of congestion control for the multimedia transmission over UMTS. We analyze TCP friendly multicast congestion control (TFMCC) over UMTS and generalize it to different multimedia classes [5][6]. We design a novel mechanism for congestion control that is Content Sensitive TCP Friendly Multicast Congestion Control (CSTFMCC). We perform a little modification in UMTS network and the packet field. At various level of network we provide the control mechanism that prevents the network from the congestion. Multimedia Class I traffic includes video and audio traffic from users equipped with an adjustable rate. Class II traffic includes non-real time data traffic such as e-mail, file transfer and web browsing traffic. These two classes contain different multimedia traffic that is more delay sensitive or less delay sensitive. So class I traffic support the real time applications and more delay sensitive. Due to congestion, if any loss of the packet or the delay between the packets can reduce the quality of received video/audio. Whether in class II traffic, if congestion occur it is acceptable to buffer non-real time data at a network node or at the user station and transmit them at a slower rate. In a large multicast group, there will usually be at least one receiver that has experienced a recent packet

loss. If the congestion control mechanisms require that the sender reduces its sending rate in response to each loss, as in TCP, then there is little potential for the construction of scalable multicast congestion control.

In wireless communication systems like UMTS, the packet loss may not mean network congestion. The quality of wireless link may be degrading due to signal fading. During a fading period, the bit error rate of wireless link may become very high but after that period the wireless link is expected to recover. TFMCC uses a feed back scheme which allows the receiver to calculating the slowest transmission rate to always reach the sender.

A. Sender end: Fragmentation of data packets perform at the sender end. The sender fragment the data packet with on bit of priority. There are two parts of the data (i) packet header and (ii) payload. The size of header part change by one bit shows the priority of packet. So there is only one bit modification perform in the size of data packet and it increases the speed of multimedia packets.

B. Multimedia Packet Size: Multimedia packet size depends on the multimedia classes. The proposed scheme redesigns the multimedia packets. It increases the multimedia packet size by one bit. This bit shows the priority of multimedia packets. The highest priority packets serve first by the routers at the layer 2. So the size of the packet is increased by one bit.

C. Routing Scenario: For fast transmission of multimedia information the proposed scheme give the priority to all multimedia packets. When a user want to send multimedia data the data framing perform at the sender end. The sender constructs the frame with a priority bit. This information stored in the header of the packet for priority access to the router. Sender sends the packets towards its destination. Multimedia packets reach at the network. At layer two the router checks the destination address and priority bit of the packet. If a higher priority packet arrives then router serves first to the packet which contains a highest priority. This increases the speed of multimedia packets and decreases the congestion in the networks.

For implementing this scenario the changes perform in the size of packets and in routers. Following algorithm shows the scenario for routing the various packets according to priority. Fig. 3 shows the algorithm for the proposed model.

```
If (incoming request for higher priority packets)
  If (there is a free channel) then
    allocate the free channel
  else
    If (lower priority packets)
      put in a buffer
      If (there is free channel again)
        allocate the free channel to lower
        priority packets
      else
        Ignore request
      endif
    else
      Ignore request
    endif
  endif
endif
End.
```

Fig.3 Algorithm For proposed model

Fig. 4 shows the flow chart for the proposed model. Flow chart shows the arrival of packet and priority check by the routers at layer2. According to this priority the packets is being processed.

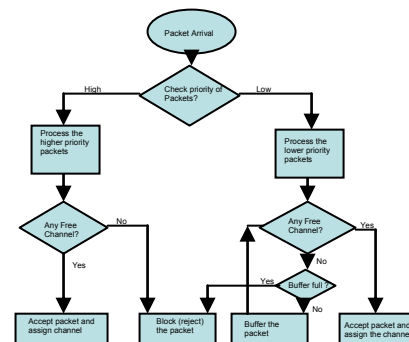


Fig. 4 Flow Chart for the proposed model

Receiving End: At the receiving end defragmentation perform. The receiving data packet reaches at the destination and multimedia information is available for the user respectively

A. Simulation Platform

The simulation that we use for this is EURANE (NS-2 Extension)[16]. Following fig. 3 shows the simulation topology to increase the multimedia quality. This paper focuses on the problem of evaluating the subjective video quality and presents the quality estimation tool that we employed. a performance evaluation study done with the well known ns-2 network simulator

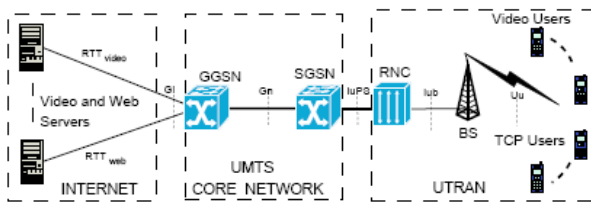


Fig:3 Simulation Topology

the video packet trace file is fed to the ns-2 simulator (compiled with the EURANE extensions). This trace file serves as a traffic generator during the simulation. A simulation script allows defining the particular scenario under consideration (network topology, simulation parameters, and so on). When the simulation is run, an output trace file is produced which contains delay- and loss-related information for every video packet sent by the (simulated) video server.

Conclusion

This paper represents to design a multicast group congestion control mechanism over the UMTS networks. This mechanism is content sensitive and optimized for multimedia traffic. This paper performs some changes in existing UMTS architecture and the packet field of data packet. User equipment (UE) has some additional function to detect the multimedia class type. The additional functionality of UE and the field of data packet has main target to remove the packet loss and congestion.

References

- [1] Holma. H. and Toskala, A. **"WCDMA for UMTS: Radio Access for Third Generation Mobile Communications"** (3rd Edition), Wiley, 2004.
- [2] Antonios Alexiou, Dimitrios Antonellis and Chritos Bouras **"A Study of Multicast Congestion Control for UMTS"** proc Int. J. Commun. Sys. (2009)
- [3] Minghua Chen and Avideh Zakhor **"Rate Control for Streaming Video over wireless"** proc of IEEE INFOCOM 2004 Hong Kong China
- [4] J. Widmer, R. Denda, M. Mauve **"A Survey on TCP-Friendly Congestion Control"** IEEE Network, 15(3), May-June 2001
- [5] Cui-Quing Yang and Alapati V.S. Reddy **"A Taxonomy for congestion control algorithms in packet switching networks"** proc IEEE 1995
- [6] Ljiljana Trajkovic and S. Jamaloddin Golestani **"Congestion Control for Multimedia Services"** proc of IEEE INFOCOM 1992
- [7] Antonios Alexiou, Dimitrios Antonellis and Chritos Bouras **"Equation Based Congestion Control for Video Transmission Over WCDMA Networks"** proc of IEEE AINA'06 Vienna, Austria, pp .445-450.
- [8] Minghua Chen and Avideh Zakhor **"Rate Control for Streaming Video over Wireless"** proc of IEEE INFOCOM 2004
- [10] Kamal Deep Singh*, Árpád Huszák., David Ros§, César Viho* and Gábor Jeney. **"Congestion Control and Adaptive Retransmission for Multimedia Streaming over Wireless Networks"** The Second International Conference on Next Generation Mobile Applications, Services, and Technologies
- [11] Kamal Deep Singh*, Julio Orozco†, David Ros‡ and Gerardo Rubino **"Improving perceived streaming-video quality in High Speed Downlink Packet Access"** proc of IEEE IEEE "GLOBECOM" 2008
- [12] S. Floyd and K. Fall. **Promoting the Use of End-to-end Congestion Control in the Internet.** IEEE/ACM Transactions on Networking, Aug. 1999.
- [13] R. Jain, K. Ramakrishnan, and D. Chiu. **Congestion Avoidance in Computer Networks with a Connectionless Network Layer.** Tech. Rep. DEC-TR-506, Digital Equipment Corporation, August 1987.
- [14] S. Floyd, M. Handley, J. Padhye, and J. Widmer. **Equation based congestion control for unicast applications.**, In Proceedings of ACM SIGCOMM 2000, pages 43.56, Stockholm, Aug. 2000.
- [15] S. Floyd, E. Kohler, and J. Padhye, **Profile for Datagram Congestion Control Protocol (DCCP) Congestion Control ID 3: TCP-Friendly Rate Control (TFRC).**, RFC 4342, IETF, Mar. 2006.
- [16] M. Handley, S. Floyd, J. Padhye, and J. Widmer. **TCP Friendly Rate Control (TFRC): Protocol Specification.** Internet Standards Track RFC 3448, IETF, Jan. 2003.
- [16] Ns-2 . Network Simulator,
<http://www.isi.edu/nsnam/ns/index.html>

Authors Profile

Ms. Neetu Sharma, Reader



Biography: Mrs. Neetu Sharma obtained her Engineering degree from University of Rajasthan and Masters Degree from Rajasthan Vidyapeeth, Udiapur securing First division with honors in both. Currently she is pursuing Ph.D. (CSE) in Congestion Control in 3G from Gyanvihar University, Jaipur, India. She has been Reader and HOD of the department of CSE at Rajasthan College of Engineering for Women, Jaipur, India. She has extensively worked in various field of Computer Engineering. She has published many national papers in the reputed journals and conferences. She is an author of the book 'System Software Engineering' for B.Tech. students. She is the member of renowned societies like IEEE, IEEE computer society, ISTE and CSI also.

Dr. Vijay Rathore, Associate Professor



Biography: Dr. Vijay Singh Rathore obtained his MCA and Ph.D. (CSE) from University of Rajasthan, India. He is an Associate Professor, Shree Karni College, Jaipur, India. He has more than 10 years of industrial and teaching experience. His areas of interest are Computer Organization &

Architecture, Operating System Fundamentals, DBMS & RDBMS (Oracle, DB2, SQL Server, MS-Access, DBASE, etc.), Data Structures, Programming Languages (C, C++, Java (J2SE, J2ME, J2EE), VB, COBOL), Networking Technologies (Data Communications, Internet & Intranet, E-Commerce, Network Security, Cryptology etc.), Software Engineering, System Analysis & Design, Management Information System, Decision Support System, Artificial Intelligence, E-Governance, Computer Center Management, UNIX, etc.. He is the member of renowned society like ISTE.

Dr. Durgesh Kumar Mishra

Professor (CSE) and Dean (R&D),
Acropolis Institute of Technology and Research,
Indore, MP, India,
Ph - +91 9826047547, +91-731-4730038
Email: durgeshmishra@ieee.org

Chairman IEEE Computer Society, Bombay Chapter
Vice Chairman IEEE MP Subsection



Biography: Dr. Durgesh Kumar Mishra has received M.Tech. degree in Computer Science from DAVV, Indore in 1994 and PhD degree in Computer Engineering in 2008. Presently he is working as Professor (CSE) and Dean (R&D) in Acropolis Institute of Technology and Research, Indore, MP, India. He is having around 21 Yrs of teaching experience and more than 7 Yrs of research experience. He has completed his research work with Dr. M. Chandwani, Director, IET-DAVV Indore, MP, India in Secure Multi- Party Computation. He has published more than 60 papers in refereed International/National Journal and Conference including IEEE, ACM etc. He is a Senior Member of IEEE, Chairman of IEEE

Computer Society, Bombay Chapter, India. Dr. Mishra has delivered his tutorials in IEEE International conferences in India as well as other countries also. He is also the programme committee member of several International conferences. He visited and delivered his invited talk in Taiwan, Bangladesh, Nepal, Malaysia, Bali-Indonesia, Singapore, Sri Lanka, USA and UK etc in Secure Multi-Party Computation of Information Security. He is an author of one book also. He is also the reviewer of tree International Journal of Information Security. He is a Chief Editor of Journal of Technology and Engineering Sciences. He has been a consultant to industries and Government organization like Sale tax and Labor Department of Government of Madhya Pradesh, India.

Mr. Amit Sharma, Assistant Professor



Biography: Mr. Amit sharma obtained his MCA from University of Rajasthan and aboout to complete his M.Tech.(CSE) from Rajasthan Technical University. He is an Assistant Professor in Sri Balaji College of Engineering & Technology, Jaipur. He has more than 5 years of industrial and teaching experience. His areas of interest are Open Source, Networking, Advance Computer Architecture and Information Security. He is the member of renowned society like ISTE.

Adaptive Sub-block ARQ techniques for wireless networks.

A.N.Kemkar, Member, ISTE and Dr. T.R.Sontakke Member,ISTE

Abstract— Wireless channels are highly affected by unpredictable factors such as co channel interference, adjacent channel interference, propagation path loss, and shadowing and multi path fading. An adaptive ARQ scheme, we mean an ARQ scheme with two or more different transmission modes meant for different channel conditions, which uses some channel sensing mechanism to decide which transmission mode is used. In this paper, we propose an adaptive transmission of sub-blocks scheme, for wireless networks. As the channel becomes increasingly noisy, the data block is divided into smaller sub-blocks for transmission. Each sub-block is encoded for error control by a CRC code. The received block is checked for errors sub-block by sub-block. The propose sub-block retransmission scheme provides improved throughput over conventional ARQ schemes by retransmitting only the same number of sub-blocks in the occurrence of errors. .

Index Terms—Retransmission protocol; Hybrid ARQ,CRC

1. INTRODUCTION:

In a mobile radio channel, burst errors frequently occur because of Rayleigh fading and shadowing. In particular, for a large cell-size system with the radius of more than several km, shadowing often becomes the predominant source of burst errors. (Shown in Fig-1)

Therefore, in order to provide reliable packet data transmission in such a channel, an efficient automatic repeat request (ARQ) protocol must be employed, since data service can tolerate some

delays but is sensitive to loss and errors. Many researchers have devoted much effort to analyze the throughput for various ARQ protocols in Rayleigh fading channels. In the packet data transmission, short packets are less likely to encounter fades than long packets, but on the other hand, they are more burdened by overheads. In other words, the packet length to maximize the instantaneous throughput is closely related to the dynamic channel condition due to fading, shadowing, and propagation path loss. Therefore, if we choose the optimum packet length adaptively by estimating the channel condition, we can continuously achieve the maximum attainable throughput.

Paper is organized as follows. We start by, describes in detail about related work in section 2. A Communication system model and proposed method of adaptation in section 3. In section 4 we have presented system analysis 5. Simulation parameters and results . Followed by conclusion in section 6.

2. RELATED WORK:

A change of transmission mode can mean, for example, a change of the packet size in the SR scheme [1] or a change of the number of transmitted copies of a packet in the GBN scheme [2] or a change of the code rate in an HARQ-I scheme [3]. In these schemes, the channel sensing is usually done by observing the acknowledgements sent by the receiver to the transmitter. This can mean either estimation of error rates, as in [4], or detection of channel state changes, as in [5] and [6], which does not require as long an observation interval (OBI) as reliable error rate estimation.

In [2], an adaptive SR scheme was proposed, where the packet size used in the current transmission was selected from a finite set of values based on a long-term BER estimate. This estimate was obtained by counting the incorrectly received packets over a time interval and

□A.N.Kemkar¹, S.R.T.M.U, Nanded.
+91-9819150392, ankemkar@gmail.com
Dr.T.R.Sontakke²

Ex.Director – S.G.G.S.I.T.E.- Nanded
Principal, Sidhant college of Engineering
Pune.+91-9822392766, trsontakke@gmail.com

assuming that there can be at most one bit error in an erroneous packet. Another adaptive SR scheme with variable packet size was proposed in [5], where the *a posteriori* distribution of the BER was computed based on the number of retransmissions during the OBI, and the packet size was selected so that the expected efficiency of the protocol was maximized. In [2], Yao proposed an adaptive GBN scheme where the transmitted number the transmitted number of copies of a packet was variable.

Numerous adaptive HARQ schemes have been suggested in the literature. Typically, the code rate is varied according to the estimated channel conditions. In [6] and [7], adaptive HARQ-I schemes were studied with convolutional codes used for error correction. Finite-state Markov models were assumed for the channel. Switching between transmission modes depended on the number of erroneous blocks occurring during an OBI. A similar adaptive HARQ-I scheme with either block or convolutional codes were proposed in [3]. In [4], sequential statistical tests were applied on the acknowledgements to detect channel state changes. An adaptive HARQ-II scheme with variable packet size was proposed for wireless ATM networks in [8]. This scheme used rate compatible convolutional (RCC) codes for error correction. In [9], three different adaptive HARQ schemes are proposed using Reed-Solomon codes for error correction. Another adaptive HARQ scheme using Reed-Solomon codes with variable rate for error correction was proposed in [10]. In this scheme short-term symbol error rate was estimated by computing the bitwise modulo-2 sum of two erroneous copies of a packet. This method was originally proposed in [11].

3. A COMMUNICATION SYSTEM MODEL AND PROPOSED METHOD OF ADAPTATION:

3.1 A communication system model:

Fig. 1 shows the communication system model. In the non cellular or large cell-size system, a radio base station continuously transmits data packets to a single mobile terminal with no packet collision after the link connection is established. Table I summarizes the digital mobile communication system characteristics, where we choose a binary frequency shift keying

non coherent detection (non coherent FSK) scheme in terms of easy implementation, because it requires no complicated carrier recovery circuit.

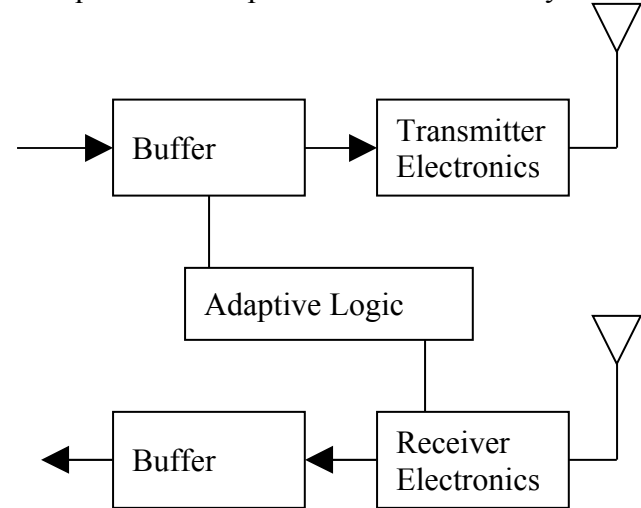


Fig.1. Communication System model.

Table 1 –Communication system parameters

Parameters	Description
Channel type	Contention Free, half duplex
Modulation /Demodulation	Binary FSK,Non coherent detection.
Packet structure	Information packet length 256,128,64,32,16 bits

3.2. Adaptation Policy

According to the variations of SNR, the receiver channel may be consider to be in one of the states at each instant ' t '. We assume that the sender knows the state at the transmission time for receiver. Let's define the *transmission status* at time t as the set of all channel state. Before transmitting, the adaptive algorithm in the sender must estimate the efficiency and packet loss rate of the ARQ/FEC protocol using all the available coding schemes as well as the ARQ protocol as a function of the transmission status. It then tries to find the protocol satisfying the desired packet loss rate. The time is divided into transmission rounds. Each transmission round corresponds to the transmission ' k ' packets in case of ARQ. A transmission round ends when the sender is informed about the reception state of the receiver. The adaptive algorithm is repeated at the end of

each transmission round. Basically, the algorithm goes through the following steps:

1. At the beginning of the algorithm, the sender determines the desired packet loss rate (SNR) of the session. It also determines the transmission status.
2. The sender estimates the packet loss rate of the ARQ protocol as well as the ARQ/FEC protocol using all the available coding schemes, based on the transmission status. It then adjusts its parameters and starts the transmission of the blocks.
3. At the end of a transmission round, the sender again determines the transmission status. It then repeats the step.

4. PERFORMANCE ANALYSIS OF THE PROPOSED SCHEME:

The performance analysis of the scheme is measured in terms of throughput of the proposed scheme. Further we show the comparison of throughput with sub block and without sub block transmission schemes with Adaptive scheme.

Expression of throughput for ARQ for present scheme:

$$\eta = \frac{K}{E[T]} \quad (1)$$

Where K = information bits in a block. $E[.]$ = Expectation of number of transmitted bits in a given block.

$$T = Mn + \sum_{i=1}^{\infty} T_i \quad (2)$$

Where M = number of sub blocks, n = number of bits in a sub block, T_i = number of transmitted bits for i^{th} transmission.

$$E[T] = Mn + \sum_{i=1}^{\infty} E[T_i] \quad (3)$$

Where $E[T]$ = Average number of transmitted bits.

Out of M sub blocks if L sub blocks are transmitted at the i^{th} retransmission, then random variable, T_i takes the value Ln , if L out of M sub-blocks are retransmitted at the i^{th} retransmission. Our algorithm compute the value of (3) to get the result from equation (1). In our analysis we have consider the variable packet size

as a retransmission units with fixed rate of data transmission. Hence we can send the packet of specific size based on the estimated signal to noise ratio (SNR). As shown in the Fig.1 the system configuration of ARQ techniques combines with adaptive packet size modulation. With an exact bit error rate equation for FSK at certain signal to noise ratio ' γ ', we can decide the value of packet size satisfying the required BER (bit error rate) Assuming that we have ' M ' different block sizes $\{L_1 L_2 L_3 L_4 \dots L_M\}$.

Let $A_i, i \in \{0, \dots, M\}$ with A_i as the threshold value of signal to noise ratio, being between the i^{th} level and $i+1$ level.

A_0 is the lowest possible signal to noise ratio and A_M is the highest possible signal to noise ratio.

5. SIMULATION RESULTS:

We evaluate the performance of the proposed scheme implemented with Matlab. We run the simulation for three schemes i.e. with sub block transmission and without sub block transmission and adaptive. The simulation parameters are shown in the table 2.

Table 2- Simulation Parameters

Parameters	Notation	Values
Signal to Noise Ratio	γ	Varied
Threshold values of SNR	snr	8,7,6,5
Max. number of Retransmissions	-----	3
Number of sub blocks retransmitted	L	Varied
Cyclic Redundancy Check	CRC	Varied
Bit error rate	BER	Varied
Packet error rate	PER	Varied
Throughput efficiency	η	Varied
Data Rate	R	9.6kbps

Simulation runs for 5000 total blocks. Result is the average of independent experiments where each experiment uses different randomly generated uniform parameters. We use mean values which are obtained independent

experiments as a basic data to get the result. Simulation results are shown Table 3.

For the packet error rate 0.3,0.5,0.7 the throughput of the system is say η_1, η_2, η_3 .

Adaptive Scheme parameters			Throughput efficiency		
SNR	Sub block	Packet size	η_1	η_2	η_3
8	2	256	0.92	0.84	0.82
7	4	128	0.93	0.85	0.83
6	8	64	0.94	0.92	0.88
5	16	32	0.97	0.95	0.93
Below this	32	16	0.98	0.94	0.96

Table 3.Simulation Results of adaptive scheme

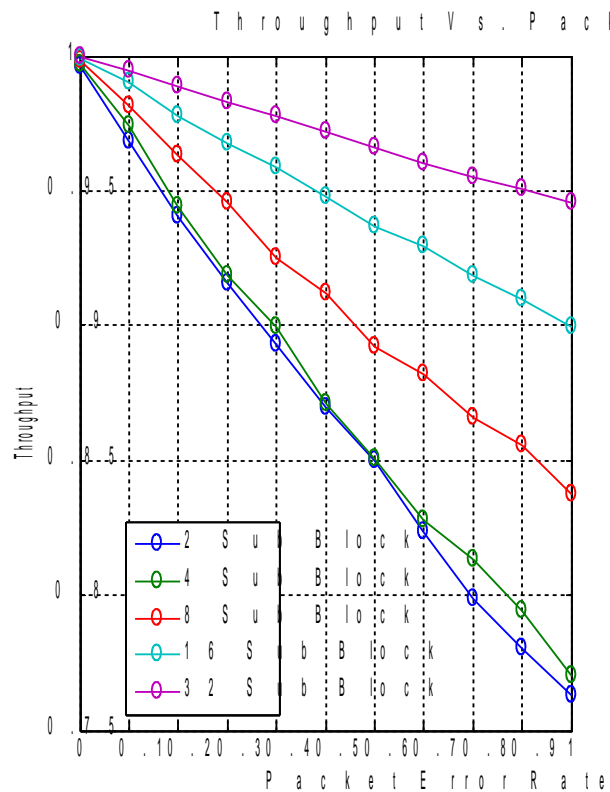


Fig. 2. Comparison as per Table -2

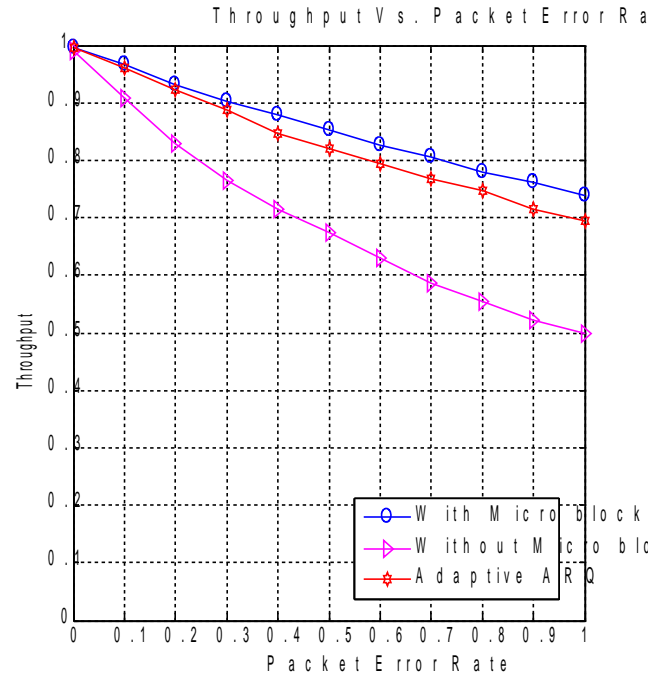


Fig. 3. Comparison of three scheme

6.CONCLUSION: Table -2 and Fig.2 shows the result of five modes of sub block transmission .Proposed Adaptive scheme will choose proper mode in accordance with the channel parameters, here threshold values of SNR (considered as channel sensing mechanism).From the Fig.3 the simulation results showed that the simulated throughput points for adaptive scheme follows the ideal throughput curve(with micro block transmission) very closely. The proposed Adaptive sub-block retransmission scheme improved the throughput and the reliability by using dynamically adapting the number of Sub block transmission according

to length to the varying channel packet error rate..

REFERENCES :

- [1] J.A.C. Martins and J.C. Alves. ARQ protocols with adaptive block size perform better over a wide range of bit error rates. *IEEE Transactions on Communications*, 38:737–739, June 1990.
- [2] Y.-D. Yao. An effective go-back-N ARQ scheme for variable-error-rate channels. *IEEE Transactions on Communications*, 43:20–23, January 1995.
- [3] M. Rice and S.B. Wicker. Adaptive error control for slowly varying channels. *IEEE Transactions on Communications*, 42:917–926, February/March/April 1994.
- [4] M. Rice and S.B. Wicker. A sequential scheme for adaptive error control over slowly varying channels. *IEEE Transactions on Communications*, 42:1533–1543, February/ March/April 1994.
- [5] E. Modiano. An adaptive algorithm for optimizing the packet size used in wireless ARQ protocols. *Wireless Networks*, 5:279–286, July 1999
- [6] B. Vucetic. An adaptive coding scheme for time-varying channels. *IEEE Transactions on Communications*, 39:653–663, May 1991.
- [7] B. Vucetic, D. Drajić, and D. Perisic. Algorithm for adaptive error control synthesis
- [8] I. Joe. A novel adaptive hybrid ARQ scheme for wireless ATM networks. *Wireless Networks*, 6:211–219, July 2000.
- [9] S. Choi and K.G. Shin. A class of adaptive hybrid ARQ schemes for wireless links. *IEEE Transactions on Vehicular Technology*, 50:777–790, May 2001.
- [10] H.Minn, M. Zeng, and V.K. Bhargava. On ARQ scheme with adaptive error control. *IEEE Transactions on Vehicular Technology*, 50:1426–1436, November 2001. *IEE Proceedings, Part F*, 135:85–94, February 1988
- [11] S.S. Chakraborty, M. Liinajarja, and E. Yli-Juuti. An adaptive ARQ scheme with packet combining for time varying channels. *IEEE Communications Letters*, 3:52–54, February 1999.

Trigon Based Authentication Service Creation with Globus Middleware

Ruckmani V ¹

Senior Lecturer, MCA,
Ramakrishna Engineering College,
Coimbatore, India

Anitha Kumari K ²

Lecturer, IT, ,
PSG College of Technology,
Coimbatore, India

Sudha Sadasivam G ³

Professor, CSE,
PSG College of Technology,
Coimbatore, India

Dhaarini M P ⁴

Lecturer, IT,
PSG College of Technology
Coimbatore, India .

Abstract— A Grid is built from multi-purpose protocols and interfaces that address fundamental issues as authentication, authorization, resource discovery, and resource access. Security is of utmost importance in grid computing applications as grid resources are heterogeneous, dynamic, and multi-domain. Authentication remains as the significant security challenge in grid environment. The proposed approach uses a dual authentication protocol in order to improve the authentication service in grid environment. The protocol utilizes the fundamental concepts of trigon and based on the parameters of the trigon the user authentication will be performed. In the proposed protocol, the password is interpreted and alienated into more than one unit and these units are stored in two different servers, namely, Authentication Server and Backend Server. Only when the combined authentication scheme from both the servers authenticates the user, the privilege of accessing the requested resources is obtained by the user. The main advantage of utilizing the dual authentication protocol in grid computing is that an adversary user cannot attain the access privilege by compromising a single consolidated server because of the fact that the split password is stored in different servers. Grid service is stateful and transient web service, which can be invoked by clients, and is considered to be the mainstream of future internet. The creation of Web Services standards is an industry-led initiative, with some of the emerging standards in various states of progress through the World Wide Web Consortium (W3C). To achieve reuse of behaviors of this authentication concept, operations are often grouped together to form a trigon based authentication service.

Keywords— *Trigonbasedauthentication, web services, globus.*

I. INTRODUCTION

Grid computing has emerged as a significant new field, distinguished from conventional distributed computing by its concentration on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation . Grid computing is concentrating on large-scale resource sharing and collaboration over enterprises and

virtual organizations boundaries. A VO is a dynamic group of individuals, groups, or organizations that have common rules for resource sharing [8]. Confidentiality of information in a VO Should also be ensured [28]. The necessity for secure communication between grid entities has motivated the development of the Grid Security Infrastructure (GSI). GSI provides integrity, protection, confidentiality and authentication for sensitive information transferred over the network in addition to the facilities to securely traverse the distinct organizations that are part of collaboration. Authentication is done by exchanging proxy credentials and authorization by mapping to a grid map file. Grid technologies have adopted the use of X.509 identity certificates to support user authentication. SOAP protocol [12] is used by the emerging OGSA. This necessitates for support message layer security using XML digital signature standard and the XML encryption standard [11]. Globus Toolkit [24] provides security services for authentication, authorization, management of user credentials and user information. Laccetti and G. Schmid [14] have introduced a unified approach for access control of grid resources. PKI (Public Key Infrastructure) and PMI (Privilege Management Infrastructure) infrastructures were utilized at the grid layer after authentication and authorization procedures. Czajkowski [5] have explained about agreement based grid management. Nagaratnam [18] have introduced security architecture for open grid services. H.-L. Truong[26] define a framework for monitoring and analyzing qos metrics of grid Services. The proposed work aims at authenticating the users by using trigon concept and to host this operation as a web service.

A. Globus Middleware

Globus [25] provides a software infrastructure that enables applications to handle distributed heterogeneous computing resources as a single virtual machine. Globus is constructed as a layered architecture in which high-level global services are built upon essential low-level core local services. Middleware is generally considered to be the layer

of software sandwiched between the operating system and applications, providing a variety of services required by an application to function correctly. Recently, middleware has re-emerged as a means of integrating software applications running in distributed heterogeneous environments. In a Grid, the middleware is used to hide the heterogeneous nature and provide users and applications with a homogeneous and seamless environment by providing a set of standardized interfaces to a variety of services.

B. Web Services

The term Web Services describes an important emerging distributed computing paradigm that differs from other approaches such as DCE, CORBA, and Java RMI in its focus on simple, Internet-based standards to address heterogeneous distributed computing. Web services define a technique for describing software components to be accessed, methods for accessing these components, and discovery methods that enable the identification of relevant service providers. Once a web service is created, it is advertised in a registry called UDDI (Universal Description, Discovery and Integration) [27], where it can be searched. The UDDI will provide the location to the service provider's WSDL (Web Services Description Language) [29] file that describes the methods that can be invoked and the parameters required. Messages are exchanged through the protocol SOAP (Simple Object Access Protocol) [30].

The established standards include:

SOAP (XML Protocol). SOAP provides an envelope which encapsulates XML data for transfer through the Web infrastructure (e.g. over HTTP, through caches and proxies), with a convention for Remote Procedure Calls (RPCs) and a serialization mechanism based on XML Schema data types. SOAP is being developed by W3C in cooperation with the Internet Engineering Task Force (IETF).

Web Services Description Language (WSDL). Describes a service in XML, using an XML Schema; there is also a mapping to the Resource Description Framework (RDF). In some ways WSDL is similar to an interface definition language IDL. WSDL is available as a W3C note [WSDL].

Universal Description Discovery and Integration (UDDI). This is a specification for distributed registries of web services, similar to yellow and white pages services. UDDI supports 'publish, find and bind': a service provider describes and publishes the service details to the directory; service requestors make requests to the registry to find the providers of a service; the services 'bind' using the technical details provided by UDDI. It also builds on XML and SOAP [UDDI].

Web Services have certain advantages over other technologies:

Web Services are platform-independent and language-independent, since they use standard XML languages. This means that my client program can be programmed in C++ and running under Windows, while the Web Service is programmed in Java and running under Linux.

Service Processes: This part of the architecture generally involves more than one Web service. For example, discovery belongs in this part of the architecture, since it allows us to locate one particular service from among a collection of Web services.

Service Description: One of the most interesting features of Web Services is that they are self-describing. This means that, once you've located a Web Service, you can ask it to 'describe itself' and tell you what operations it supports and how to invoke it. This is handled by the Web Services Description Language (WSDL).

Service Invocation: Invoking a Web Service (and, in general, any kind of distributed service such as a CORBA object or an Enterprise Java Bean) involves passing messages between the client and the server. SOAP (Simple Object Access Protocol) specifies how we should format requests to the server, and how the server should format its responses. In theory, we could use other service invocation languages (such as XML-RPC, or even some ad hoc XML language). However, SOAP is by far the most popular choice for Web Services.

Transport: Finally, all these messages must be transmitted somehow between the server and the client. The protocol of choice for this part of the architecture is HTTP (Hypertext Transfer Protocol), the same protocol used to access conventional web pages on the Internet. Again, in theory we could be able to use other protocols, but HTTP is currently the most used one.

C. Web Service Definition Language(WSDL)

Web Services programmers usually only have to concentrate on writing code in their favorite programming language and, in some cases, in writing WSDL. SOAP code, on the other hand, is always generated and interpreted automatically for us. Once we've reached a point where our client application needs to invoke a Web Service, we delegate that task on a piece of software called a stub. Using stubs simplifies our applications considerably. We don't have to write a complex client program that dynamically generates SOAP requests and interprets SOAP responses (and similarly for the server side of our application). We can simply concentrate on writing the client and/or server code, and leave all the dirty work to the stubs (which, again, we don't even have to write ourselves... they can be generated automatically from the WSDL description of a web service). The stubs are generally generated only once. In general, we only go through the discovery step once, then generate the stubs once (based on the WSDL of the service we've discovered) and then reuse the stubs as many times as we want (unless the maintainers of the Web service decide to

change the service's interface and, thus, its WSDL description).

II. TRIGON BASED AUTHENTICATION ARCHITECTURE

When legitimate entities (users) login, the trigon based authentication server splits the password into its components and stores the authentication information in two servers – namely authentication and backend server. Users have to register with the Authentication server, so that it can hold a part of the interpreted password with itself and another part in the Backend server. The block diagram illustrating the registration process of the users is depicted in the Figure 5. As illustrated in Figure 5, the users who require services from the VO have to register initially with the Authentication server using their username and password. The Authentication server calculates the P_i as given in (1). Along with this authentication server generates two large prime numbers, namely, a and a' , which are considered as the two sides of a trigon. It is difficult to hack the values of a and a' as they are large prime numbers (as per RSA Factoring Challenge). Here, P_i is taken as the angle between the two sides of the trigon a and a' . Now, the Authentication server can easily determine the opposite side of the angle P_i , termed as a'' .

With these trigon parameters, α , $V_{aa'}$ and $P_{aa'}$ are found as

$$V_{aa'} = a - a' \quad (1)$$

$$P_{aa'} = a * a' \quad (2)$$

$$\alpha = 2P_{aa'} - a'^2 \quad (3)$$

where, a , a' and a'' are the three sides of trigon. α is a strengthening parameter used as the index. $V_{aa'}$ and $P_{aa'}$ are the variance and the product of the sides a and a' respectively. With the parameters a , a' and a'' as the sides of trigon and P_i be the angle between the sides a and a' the generated trigon will be assumed. After the calculation of α , $V_{aa'}$ and $P_{aa'}$, the authentication server stores the α value and its corresponding username in a database and forwards $V_{aa'}$ and $P_{aa'}$ to the Backend server along with the username. Hence, the password is interpreted and alienated into two units and stored in two separate server. The authentication procedure is based on the fundamental concepts of a trigon. Initially, the user who wants the services of VO has to login to the Authentication server using the username and password. Here, u_i and p_{wi} refers to username and password of i th user. The Authentication server calculates the Password index (P_i) from the password as

$$P_i = \begin{cases} P_{AI(i) / 10^{\text{pow } n-2}} & ; \text{ if } P_{AI(i)} \geq 180 \\ P_{AI(i) / 10^{\text{pow } n-3}} & ; \text{ else} \end{cases} \quad (4)$$

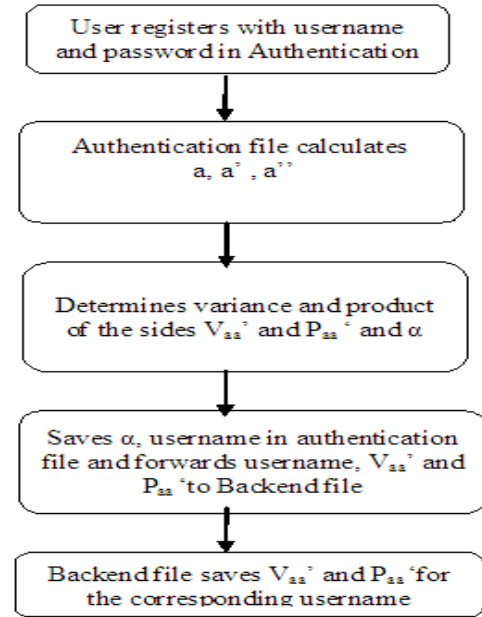


Fig 1 Flow Diagram

In (4), PAI is the ASCII-interpreted value of the given password p_{wi} , n is the total number of digits in PAI and $PAI(j)$ represents the first j digits of PAI. The PAI can be calculated by the following steps.

Change the p_{wi} into its corresponding ASCII value.

Calculate the three-fourth of total digits of the ASCII value modulo 180, which results the first three digits of PAI.

Append the remaining one-fourth of the ASCII digits to PAI.

Then, from P_i the Authentication Server determines the Authentication index (AI) for u_i as

$$AI(i) = P_i / 2 \quad (5)$$

Then, the Authentication Server searches for the username index α_i for the corresponding u_i which has already been stored in the server database during the process of the registration. Subsequently, α_i is sent to the backend server along with u_i . When the Backend server receives the index α_i and the username from the Authentication server, it searches for $V_{aa'}$ and $P_{aa'}$ the Variance and the product of the sides a and a' respectively, which have been saved in the backend server database during the process of registration. From these values,

the Backend server calculates an Authentication Token AT_i and sends it to the Authentication server to authenticate the ui. The AT_i can be calculated as

$$AT(i) = \alpha_i + Vaa'I * 2Paa'I \quad (6)$$

In (6), Vaa' and Paa' are pre-calculated values computed during individual user registration. After retrieval of AT_i from the Backend server, the Authentication server authenticates the user based on the token from the Backend server and the index calculated at the Authentication server. The authentication code (or) condition which authenticates the ui is given by

$$\text{Sin AI}(i) = (1 - AT_i / 2)^{1/2} \quad (7)$$

The authentication process is performed by the authentication condition given in (7). When the condition is satisfied, the user is decided to be valid and the Server sends a token called Token for VO access to the user.

III. IMPLEMENTATION - AUTHENTICATION AS SERVICE

A service is an entity that provides some capability to its clients by exchanging messages. A service is defined by identifying sequences of specific message exchanges that cause the service to perform some operation. By thus defining these operations only in terms of message exchange, we achieve great flexibility in how services are implemented and where they may be located. A service-oriented architecture is one in which all entities are services, and thus any operation visible to the architecture is the result of message exchange.

Prerequisites are:

build.xml
globus-build-service.sh

1. Creation of auth.wsdl File

```
<?xml version="1.0" encoding="UTF-8"?>
<definitions name="auth"
targetNamespace="http://www.globus.org/namespaces/add/h
ello_instance"
xmlns="http://schemas.xmlsoap.org/wsdl/"
```

```
xmlns:tns="http://www.globus.org/namespaces/add/hello_in
stance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<types>
<xsd:schema
targetNamespace="http://www.globus.org/namespaces/add/h
ello_instance"
xmlns:tns="http://www.globus.org/namespaces/add/hello_in
stance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<xsd:element name="addition">
<xsd:complexType>
<xsd:sequence>
<xsd:element name="input1" type="xsd:int"/>
<xsd:element name="input2" type="xsd:int"/>
</xsd:sequence>
</xsd:complexType>
</xsd:element>
<xsd:element name="response" type="xsd:string"/>
<xsd:element
name="additionrequest"
type="xsd:string"/>
</xsd:schema>
</types>
<message name="AddInputMessage">
<part
name="parameters"
element="tns:additionrequest"/>
</message>
<message name="AddOutputMessage">
<part name="resp" element="tns:response"/>
</message>
<portType name="authPortType">
<operation name="addition">
<input message="tns:AddInputMessage"/>
<output message="tns:AddOutputMessage"/>
</operation>
</portType>
</definitions>
```

2. Create namespace2package.mappings for mapping instances, bindings and services.

3. Write Implementation program .

4.Create deploy-server.wsdd

```
[globus@g20 service]$ vi deploy-server.wsdd
```

5.Create deploy-jndi-config.xml

```
[globus@g20 service]$ vi deploy-jndi-config.xml
```

6.Build the service

```
[globus@g20 ~example]$ sh globus-build-service.sh -d  
org/add/service/ -s schema/add/hello.wsdl
```

7. After the successful building Grid Archive(GAR) file has been created. Now we have to deploy the GAR file using globus-deploy-gar command.

```
[globus@g20~example]$ globus-deploy-gar  
org_add_service.gar
```

8. After successful deployment of the GAR file start the globus container.

```
[globus@g20 ~]$ globus-start-container
```

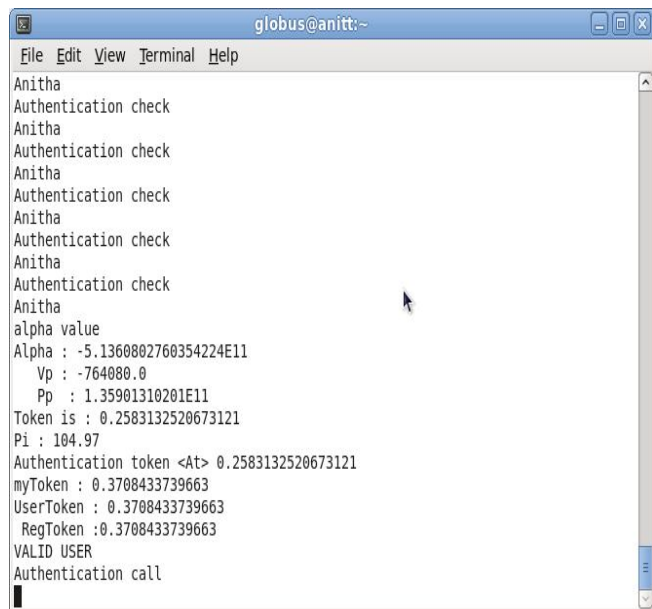


Fig 2. GUI – Server side authentication

```
[35]:  
http://192.168.100.3:8443/wsrf/services/DefaultTriggerService
```

```
[36]:  
http://192.168.100.3:8443/wsrf/services/TrigonBasedAuthenticationService
```

```
[37]:  
http://192.168.100.3:8443/wsrf/services/TriggerService
```

```
[38]:  
http://192.168.100.3:8443/wsrf/services/gsi/AuthenticationService
```

```
[39]:  
http://192.168.100.3:8443/wsrf/services/TestRPCService
```

```
[40]:  
http://192.168.100.3:8443/wsrf/services/ManagedMultiJobService
```

9. Write Client Program for authentication.

10. Before running the compiler, make sure to run the following:

```
source $GLOBUS_LOCATION/etc/globus-devel-env.sh
```

The globus-devel-env.sh script takes care of putting all the Globus libraries into your CLASSPATH.

```
[globus@gcluster example]$ source /usr/local/globus-4.0.7/etc/globus-devel-env.sh
```

```
11.[globus@gcluster example]$ javac  
org/add/client/Client.java
```

```
12. [globus@gcluster example]$ java  
org/add/client/Client
```

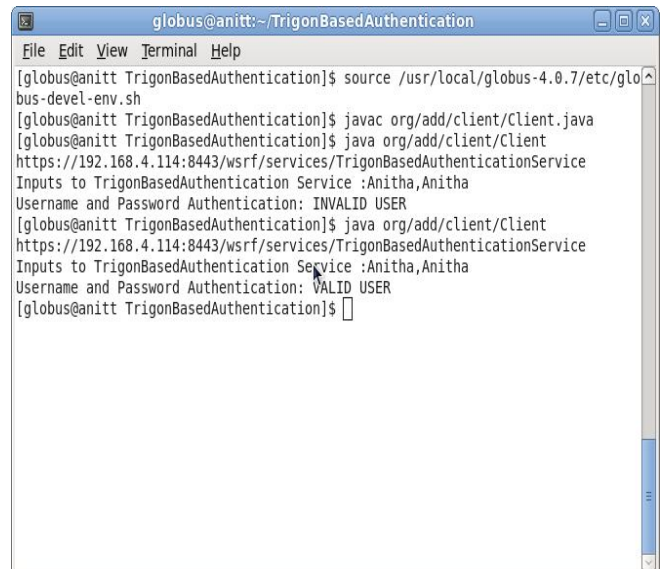


Fig 3. GUI – Valid User

The user is invalid since the username and password haven't stored in authentication and backend file. First time when user login his/her trigon value of the password gets stored in the respective files. Next time when they login they gets authenticated and token gets issued .

IV. CONCLUSION

The Internet is a reasonable model for the Grid, providing both an early version of its services and a platform from which to evolve. The authentication protocol, proposed here, enhanced the grid security as the authentication mechanism utilized two servers for authentication. This simple trigon concept utilization in the authentication protocol introduced a novel and revolutionary idea in the authentication mechanism as well as in grid environment. The implementation of our dual authentication protocol showed its effective performance in pinpointing the adversaries and paving the way to valid users for access with the VO for resource sharing. So by hosting this authentication as a service it make the grid environment more secure. In future these services will be located by type instead by names.

ACKNOWLEDGMENT

Our thanks to Dr.R.Rudramoorthy,Principal,PSG College of Technology and Mr.K.Chidambaram, Director, Grid and Cloud systems group, Yahoo software development, India Private Limited for their support. This project is carried out in Grid and Cloud lab, PSG College of Technology.

REFERENCES

- [1] S. Albayrak, S. Kaiser, and J. Stender. Advanced grid management software for seamless services. *Multiagent Grid Syst.*, 1(4):263-270, 2005.
- [2] Andrew Johnson, Carl Kesselman, Jason Leigh, and Steven Tuecke, Application Experiences with the Globus Toolkit, Seventh IEEE International Symposium on High Performance Distributed Computing (HPDC-7 ... T).
- [3] Antonioletti, M., and Jackson, M., OGSA-DAI Product Overview, 2003. Available at www.ogsa-dai.org.uk/downloads/docs/OGSA-DAI-USER-M3-PRODUCTOVERVIEW.pdf.
- [4] Christensen, E., Curbera, F., Meredith, G., and Weerawarana, S., Web Services Description Language (WSDL) 1.1, W3C, Note 15, 2001. Available at www.w3.org/TR/wsdl.
- [5] K. Czajkowski, A. Dan, J. Rofrano, S. Tuecke, and M. Xu. Agreement-based grid service management (ogsiagreement), Proceedings of 6th IEEE/ACM International Workshop on Grid Computing (Grid2005).
- [6] Della-Libera, G. et al. (2002) Web Services, Secure Conversation Languages (WS-Secure Conversation). Version 1.0, available at <http://msdn.microsoft.com/library/default.asp?url=/lib/enus/dnglobspec/html/wssecureconversation.asp> (accessed on 2002).
- [7] Eastlake, D. and Reagle, J. (Eds.) (2002) XML Encryption Syntax and Processing. W3C Recommendation, available at <http://www.w3.org/TR/xmlenc-core/> (accessed on December 2002).
- [8] Foster, I., Kesselman, C. and Tuecke, S., "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", *International Journal of High Performance Computing Applications*, vol. 15, no.3, pp. 200-222, 2001.
- [9] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The physiology of the grid: An open grid services architecture for distributed systems integration, 2002.
- [10] I. foster, Argonne National Laboratory and the University of Chicago ,carl kesselman, Information Sciences Institute, University of Southern California "THE GRID 2 blueprint for a new computing infrastructure",2004
- [11] Graham, S., Simeonov, S., Boubez, T. etc, "Building Web Service with Java: Making Sense of XML, SOAP, WSDL and UDDI", Indianapolis, IN: Sams Publishing, 2002.
- [12] Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J-J. and Nielsen, H.F. (2003) SOAP Version 1.2 Part 1: Messaging Framework. W3C Recommendation, Available at <http://www.w3.org/TR/soap12-part1/> (accessed on June 2003).
- [13] Grid Archive Creation: <http://gdp.globus.org/gt4-tutorial/multiplehtml/ch08s02.html>.
- [14] G. Laccetti and G. Schmid, "A framework model for grid security", *Future Generation Computer Systems*, vol. 23, no. 5, pp.702-713, June 2007.
- [15] Li, Y., Jin, H., Zou, D., Chen, J. and Han, Z. (2007) 'A scalable service scheme for secure group communication in grid', 31st Annual International Computer Software and Applications Conference (COMPSAC 2007).
- [16] Li, Y., Jin, H., Zou, D., Liu, S. and Han, Z. (2008) 'An authenticated encryption mechanism for secure group communication in grid', 2008 International Conference on Internet Computing in Science and Engineering.
- [17] Li Hongweia, Sun Shixina and Yang Haomiaoa, "Identity-based authentication protocol for grid", *Journal of Systems Engineering and Electronics*, Vol. 19, no. 4, pp.860-865, August 2008.
- [18] Nagaratnam, N., Janson, P., Dayka, J., Nadalin, A., Siebenlist, F., Welch, V., Foster, I., and Tuecke, S., Security architecture for Open Grid Services, 2002. Available at www.globus.org/ogsa/Security/draft-ggf-ogsa-sec-arch-01.pdf.
- [19] Open Grid Services Architecture Data Access and Integration (OGSA-DAI) Project: www.ogsa-dai.org.uk.
- [20] K. Rochford, B. A. Coghlan, and J. Walsh. An agent-based approach to grid service monitoring. In *Proc. International Symposium on Parallel and Distributed Computing (ISPDC 2006)*, July July, 2006.
- [21] Shirasuna, S., Nakada, H., Matsuoka, S., and Sekiguchi, S., Evaluating Web services based implementations of GridRPC, in 11th IEEE International Symposium on High Performance Distributed Computing, Edinburgh, Scotland. IEEE Computer Society Press, Los Alamitos, CA, 2001.
- [22] Siebenlist, F., Welch, V., Tuecke, S., Foster, I., Nagaratnam, N., Janson, P., Dayka, J., and Nadalin, A., Roadmap towards a secure OGSA. Global Grid Forum, draft, 2002.
- [23] Shengxian Luo, Xiaochuan Peng, Shengbo Fan Peiyu Zhang, Study on Computing Grid Distributed Middleware and Its Application, International Forum on Information Technology and Application, 2009.
- [24] Thatte, S., XLANG, Web services for business process design Web site, Microsoft Corporation. www.getdotnet.com/team/xml_wsspecs/xlang-c/default.htm..
- [25] Ian Foster ,Carl Kesselman, The Globus Project: A Status Report, 5th IEEE Symp. on High. Performance Distributed Computing.
- [26] H.-L. Truong, R. Samborski, and T. Fahringer. Towards a framework for monitoring and analyzing qos metrics of grid services. In *E-SCIENCE '06: Proceedings of the Second IEEE International*

Conference on e-Science and Grid Computing, page 65, Washington, DC, USA, 2006. IEEE Computer Society.

- [27] UDDI. The UDDI technical white paper, <http://www.uddi.org/>, 2000.
- [28] Von Welch, Frank Siebenlist, Ian Foster, John Bresnahan, Karl Czajkowski, Jarek Gawor, Carl Kesselman, Sam Meder, Laura Pearlman and Steven Tuecke, "Security for Grid Services", in proceedings of the 12th IEEE International Symposium on High Performance Distributed Computing, pp.48- 57, June 2003.
- [29] W3C Note "Web Services Definition Language (WSDL) 1.1", <http://www.w3.org/TR/WSDL>.
- [30] W3C Note "Simple Object Access Protocol(SOAP) 1.1", <http://www.w3.org/TR/WSDL>.

Cloud Computing. She has presented 1 paper in National Conference. She is the Best Outgoing Student in MTech 2010-2011 in PSG College of Technology. You may contact her at dhaarinimp@gmail.com

AUTHORS PROFILE



Dr G Sudha Sadasivam is working as a Professor in Department of Computer Science and Engineering in PSG College of Technology, India. Her areas of interest include, Distributed Systems, Distributed Object Technology, Grid and Cloud Computing. She has published 20 papers in referred journals and 32 papers in National and International Conferences. She has coordinated two AICTE – RPS projects in

Distributed and Grid Computing areas. She is also the coordinator for PSG-Yahoo Research on Grid and Cloud computing. You may contact her at sudhasadhasivam@yahoo.com



V Ruckmani received B. Sc, MCA and M. Phil degrees from the department of computer science, Bharathiar University, India in 1994, 1997 and 2003 respectively. She is currently pursuing the Ph. D degree, working closely with Prof. G. Sudha Sadasivam. From 1997 to 2000 she worked at PSG College of Arts and Science in the department of Computer Science. Since December 2000 she is working as a senior lecturer in Department of Computer Applications

in Sri Ramakrishna Engineering College, India. She works in the field of Grid Computing specializing in the area of security. You may contact her at ruckmaniv@yahoo.com



K Anitha Kumari received BE(CSE) from Department of Computer Science ,Avinashilingam Deemed University and ME(SE) from Department of Computer Science ,Anna University. She is working as a Lecturer in Department of Information Technology in PSG College of Technology, India. Her areas of interest include Grid and Cloud Computing. She has published 1 paper in referred international journal and 5 papers in National and

International Conferences. She is the Best Outgoing Student in ME 2010-2011 in PSG College of Technology. She awarded Gold Medal in BE(CSE) in Avinashilingam Deemed University . You may contact her at anitha.psgsoft@gmail.com



M P Dhaarini received BTech(IT) from Department of Information Technology ,Anna University and MTech(IT) from Department of Information Technology, Anna University. She is working as a Lecturer in Department of Information Technology in PSG College of Technology, India. Her areas of interest include

Performance Evaluation of Speaker Identification for Partial Coefficients of Transformed Full, Block and Row Mean of Speech Spectrogram using DCT, WALSH and HAAR

Dr. H. B. Kekre

Senior Professor,
MPSTME, SVKM's NMIMS
University
Mumbai, 400-056, India

Dr. Tanuja K. Sarode

Assistant Professor,
Thadomal Shahani Engg.
College, Bandra (W),
Mumbai, 400-050, India

Shachi J. Natu

Lecturer,
Thadomal Shahani Engg.
College, Bandra (W),
Mumbai, 400-050, India

Prachi J. Natu

Assistant Professor,
GVAIET, Shelu
Karjat 410201,
India

Abstract- In this paper an attempt has been made to provide simple techniques for speaker identification using transforms such as DCT, WALSH and HAAR alongwith the use of spectrograms instead of raw speech waves. Spectrograms form a image database here. This image database is then subjected to different transformation techniques applied in different ways such as on full image, on image blocks and on Row Mean of an image and image blocks. In each method, results have been observed for partial feature vectors of image. From the results it has been observed that, transform on image block is better than transform on full image in terms of identification rate and computational complexity. Further, increase in identification rate and decrease in computations has been observed when transforms are applied on Row Mean of an image and image blocks. Use of partial feature vector further reduces the number of comparisons needed for finding the most appropriate match.

Keywords- Speaker Identification, DCT, WALSH, HAAR, Image blocks, Row Mean, Partial feature vector.

I. INTRODUCTION

To provide security in a multiuser environment, it has become crucial to identify users and to grant access only to those users who are authorized. Apart from the traditional login and password method, use of biometric technology for the authentication of users is becoming more and more popular nowadays. Biometrics comprises methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits. Biometric characteristics can be divided in two main classes: Physiological which are related to the shape of the body. Examples include fingerprint, face recognition, DNA, hand and palm geometry, iris recognition etc. Behavioral, which are related to the behavior of a person. Examples include typing rhythm, gait and voice. Techniques like face recognition, fingerprint recognition and retinal blood vessel patterns have their own drawbacks. To identify an

individual by these methods, he/she should be willing to undergo the tests and should not get upset by these procedures. Speaker recognition allows non-intrusive monitoring and also achieves high accuracy rates which conform to most security requirements. Speaker recognition is the process of automatically recognizing who is speaking based on some unique characteristics present in speaker's voice [2]. There are two major applications of speaker recognition technologies and methodologies: speaker identification and speaker verification.

In the speaker identification task, a speech utterance from an unknown speaker is analyzed and compared with speech models of known speakers. The unknown speaker is identified as the speaker whose model best matches the input utterance. In speaker verification, an identity is claimed by an unknown speaker, and an utterance of this unknown speaker is compared with a model for the speaker whose identity is being claimed. If the match is good enough, that is, above a threshold, the identity claim is accepted. The fundamental difference between identification and verification is the number of decision alternatives [3]. In identification, the number of decision alternatives is equal to the size of the population, whereas in verification there are only two choices, acceptance or rejection, regardless of the population size. Therefore, speaker identification performance decreases as the size of the population increases, whereas speaker verification performance approaches a constant, independent of the size of the population, unless the distribution of physical characteristics of speakers is extremely biased.

Speaker identification can be further categorized into text-dependent and text independent speaker identification based on the relevance to speech contents [2, 4].

Text Dependent Speaker Identification requires the speaker saying exactly the enrolled or given password/speech. Text Independent Speaker Identification is a process of verifying the identity without constraint on the speech content. It has no

advance knowledge of the speaker's utterance and is more flexible in situation where the individuals submitting the sample may be unaware of the collection or unwilling to cooperate, which presents more difficult challenge.

Compared to Text Dependent Speaker Identification, Text Independent Speaker Identification is more convenient because the user can speak freely to the system. However, it requires longer training and testing utterances to achieve good performance. Text Independent Speaker Identification is more difficult problem as compared to Text Dependent Speaker Identification because the recognition system must be prepared for an arbitrary input text.

Speaker Identification task can be further classified into closed set and open set identification.

In closed set problem, from N known speakers, the speaker whose reference template has the maximum degree of similarity with the template of input speech sample of unknown speaker is obtained. This unknown speaker is assumed to be one of the given set of speakers. Thus in closed set problem, system makes a forced decision by choosing the best matching speaker from the speaker database.

In the open set text dependent speaker identification, matching reference template for an unknown speaker's speech sample may not exist. So the system must have a predefined tolerance level such that the similarity degree between the unknown speaker and the best matching speaker is within this tolerance.

In the proposed method, speaker identification is carried out with spectrograms and transformation techniques such as DCT, WALSH and HAAR [15-18]. Thus an attempt is made to formulate a digital signal processing problem into pattern recognition of images.

The rest of the paper is organized as follows: in section II we present related work carried out in the field of speaker identification. In section III our proposed approach is presented. Section IV elaborates the experiment conducted and results obtained. Analysis of computational complexity is presented in section V. Conclusion has been outlined in section VI.

II. RELATED WORK

All speaker recognition systems at the highest level contain two modules, feature extraction and feature matching.

Feature extraction is the process of extracting subset of features from voice data that can later be used to identify the speaker. The basic idea behind the feature extraction is that the entire feature set is not always necessary for the identification process. Feature matching is the actual procedure of identifying the speaker by comparing the extracted voice data with a database of known speakers and based on this suitable decision is made.

There are many techniques used to parametrically represent a voice signal for speaker recognition task. One of the most popular among them is Mel-Frequency Cepstrum Coefficients (MFCC) [1].

The MFCC parameter as proposed by Davis and Mermelstein [5] describes the energy distribution of speech signal in a frequency field. Wang Yutai et. al. [6] has proposed a speaker recognition system based on dynamic MFCC parameters. This technique combines the speaker information obtained by MFCC with the pitch to dynamically construct a set of the Mel-filters. These Mel-filters are further used to extract the dynamic MFCC parameters which represent characteristics of speaker's identity.

Sleit, Serhan and Nemir [7] have proposed a histogram based speaker identification technique which uses a reduced set of features generated using MFCC method. For these features, histograms are created using predefined interval length. These histograms are generated first for all data in feature set for every speaker. In second approach, histograms are generated for each feature column in feature set of each speaker.

Another widely used method for feature extraction is use of linear Prediction Coefficients (LPC). LPCs capture the information about short time spectral envelope of speech. LPCs represent important speech characteristics such as formant speech frequency and bandwidth [8].

Vector Quantization (VQ) is yet another approach of feature extraction [19-22]. In Vector Quantization based speaker recognition systems; each speaker is characterized with several prototypes known as code vectors [9]. Speaker recognition based on non-parametric vector quantization was proposed by Pati and Prasanna [10]. Speech is produced due to excitation of vocal tract. Therefore in this approach, excitation information can be captured using LP analysis of speech signal and is called as LP residual. This LP residual is further subjected to non-parametric Vector Quantization to generate codebooks of sufficiently large size. Combining nonparametric Vector Quantization on excitation information with vocal tract information obtained by MFCC was also introduced by them.

III. PROPOSED METHODS

In the proposed methods, first we converted the speech samples collected from various speakers into spectrograms [11]. Spectrograms were created using Short Time Fourier Transfer method as discussed below:

In the approach using STFT, digitally sampled data are divided into chunks of specific size say 128, 256 etc. which usually overlap. Fourier transform is then obtained to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image, which is a measurement of magnitude versus frequency for a specific moment in time.

Thus we converted the speech database into image database. Different transformation techniques such as Discrete Cosine Transform [12], WALSH transform and HAAR transform are then applied to these images in three different ways to obtain their feature vectors.

1. Transform on full image
2. Transform on image blocks obtained by dividing an image into four equal and non-overlapping blocks

3. Transform on Row Mean of an image and on Row Mean of image blocks.

From these feature vectors, again identification rate is obtained for various portions selected from the feature vector i.e. for partial feature vector [15, 23, 24]. Two different sets of database were generated. First set, containing 60% of the total images as trainee images and 40% of the total images as test images. Second set, containing 80% of the images as trainee images and 20% of the total images as test images. Euclidean distance between test image and trainee image is used as a measure of similarity. Euclidean distance between the points $X(X_1, X_2, \text{etc.})$ and point $Y(Y_1, Y_2, \text{etc.})$ is calculated using the formula shown in equation. (1).

$$D = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Smallest Euclidean distance between test image and trainee image means the most probable match of speaker. Algorithms for transformation technique on full image and transformation techniques on image blocks are given below.

A. Transformation techniques on full image[27, 28]:

In the first method 2-D DCT / WALSH / HAAR is applied on the full image resized to 256*256. Further, instead of full feature vector of an image only some portion of feature vector i.e. partial feature vector is selected for identification purpose. This selection of feature vector is illustrated in Fig. 1 and it is based on the number of rows and columns that have been selected from the feature vector of an image. For example, initially first full feature vector (i.e. 256*256) has been selected and then partial feature vectors of size 192*192, 128*128, 64*64, 32*32, 20*20 and 16*16 were selected from the feature vector. For these different sizes, identification rate was obtained.

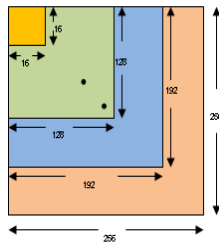


Fig. 1: Selection of partial feature vector

Algorithm for this method is as follows:

- Step 1.** For each trainee image in the database, resize an image to size 256*256.
- Step 2.** Apply the transformation technique (DCT / WALSH / HAAR) on resized image to obtain its feature vector.
- Step 3.** Save these feature vectors for further comparison.
- Step 4.** For each test image in the database, resize an image to size 256*256.
- Step 5.** Apply the transformation technique (DCT / WALSH / HAAR) on resized image to obtain its feature vector.
- Step 6.** Save these feature vectors for further comparison.
- Step 7.** Calculate the Euclidean distance between feature vectors of each test image with each trainee image corresponding to the same sentence.
- Step 8.** Select the trainee image which has smallest Euclidean distance with the test image and declare the speaker corresponding to this trainee image as the identified speaker.

Repeat Step 7 and Step 8 for partial feature vector obtained from the full feature vector.

B. Transformation technique on image blocks[27, 29]:

In this second method, resized image of size 256*256 is divided into four equal parts as shown in Fig. 2 and then 2-D DCT / WALSH / HAAR is applied to each part.

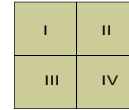


Fig. 2: Image divided into four equal non-overlapping blocks

Thus when $N*N$ image is divided into four equal and non-overlapping blocks, blocks of size $N/2*N/2$ are obtained. Feature vector of each block when appended as columns forms a feature vector of an image. Thus size of feature vector of an image in this case is of 128*512. Again Euclidean distance is used as a measure of similarity. Here also using partial feature vectors, identification rate has been obtained. Partial feature vectors of size 96*384, 64*256, 32*128, 16*64 and 8*32 have been selected to find identification rate. Detailed steps are explained in algorithm given below:

- Step 1.** For each trainee image in the database, resize an image to size 256*256.
- Step 2.** Divide the image into four equal and non-overlapping blocks as explained in Fig. 2.
- Step 3.** Apply transformation technique (DCT/ WALSH /HAAR) on each block obtained in Step 2.
- Step 4.** Append the feature vectors of each block one after the other to get feature vector of an image.
- Step 5.** For each test image in the database, resize an image to size 256*256.
- Step 6.** Divide the image into four equal and non-overlapping blocks as shown in Fig. 2.
- Step 7.** Apply transformation technique (DCT /WALSH /HAAR) on each block obtained in Step 6.
- Step 8.** Append the feature vectors of each block one after the other to get feature vector of an image.

- Step 9.** Calculate the Euclidean distance of each test image with each trainee image corresponding to the same sentence.
- Step 10.** Select the trainee image which has smallest Euclidean distance with the test image and declare the speaker corresponding to this trainee image as the identified speaker.

Repeat Step 9 and Step 10 for partial feature vectors selected from feature vector obtained in Step 4 and Step 8. Selection of partial feature vector is similar to the one shown in Fig. 1. But in this method, size of feature vector is 128*512, 96*384, 64*256, 32*128, 16*64 and 8*32.

C. Transformation techniques on Row Mean [16-18] of an image and on Row Mean of image blocks [27, 29]:

In this approach, Row Mean of an image is calculated. Row mean is nothing but an average of pixel values of an image along each row. Fig. 3 shows how the Row Mean of an image is obtained.

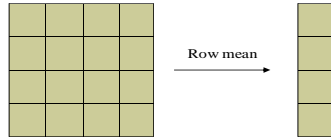


Fig. 3: Row Mean of an image

1-D DCT / WALSH / HAAR is then applied on this Row mean of an image to its feature vector and Euclidean distance is used as measure of similarity to identify speaker. Detail algorithm is given below:

- Step 1:** For each trainee image in the database, resize an image to size 256*256.
- Step 2:** Calculate Row Mean of an image as shown in Fig. 3.
- Step 3:** Apply 1-D transformation technique (DCT / WALSH / HAAR) on Row Mean obtained in Step 2. This gives the feature vector of an image.
- Step 4:** For each test image in the database, resize an image to size 256*256.
- Step 5:** Apply 1-D transformation technique (DCT / WALSH / HAAR) on Row Mean obtained in Step 4. This gives the feature vector of an image.
- Step 6:** Calculate the Euclidean distance of each test image with each trainee image corresponding to the same sentence.
- Step 7:** Select the trainee image which has smallest Euclidean distance with the test image and declare the speaker corresponding to this trainee image as the identified speaker.

For Row Mean of image blocks, first divide the image into equal and non-overlapping blocks (of size 128*128, 64*64, 32*32, 16*16 and 8*8). Obtain Row Mean of each block as shown in Fig. 3. Transformation technique is then applied on Row Mean of each block and then combined into columns to get feature vector of an image.

IV. EXPERIMENTS AND RESULTS

Implementation for the proposed approach was done on Intel Core 2 Duo Processor, 2.0 GHz, and 3 GB of RAM. Operating System used is Windows XP and softwares used are MATLAB 7.0 and Sound forge 8.0. To study the proposed approach we recorded six distinct sentences from 30 speakers: 11 males and 19 females. These sentences are taken from VidTIMIT database [13] and ELSDSR database [14]. For every speaker 10 occurrences of each sentence were recorded. Recording was done at varying times. This forms the closed set for our experiment. From these speech samples spectrograms were created with window size 256 and overlap of 128. Before creation of spectrograms, DC offset present in speech samples was removed so that signals are vertically centered at 0. After removal of DC offset, speech samples were normalized with respect to amplitude to -3 dB and also with respect to time. Spectrograms generated from these speech samples form the image database for our experiment. In all we had 1800 spectrograms in our database.

From these spectrograms, two sets were created.

Set A: Contains six spectrograms as trainee images per speaker and four spectrograms as test images per speaker. So in all it contains 1080 trainee images and 720 test images.

Set B: Contains eight spectrograms as trainee images per speaker and two spectrograms as test images per speaker. So in all it contains 1440 trainee images and 360 test images.

Since our work is restricted to text dependent approach, Euclidean distance for a test image of speaker say 'x' for a particular sentence say 's1' is obtained by comparing the feature vector of that test image with the feature vectors of all the trainee images corresponding to sentence 's1'. Results are calculated for set of test images corresponding to each sentence.

A. Results for DCT/WALSH/HAAR on Full image:

1) Results for DCT on full image

Table I shows the identification rate for six sentences s1 to s6 when DCT is applied on full image in set A and partial feature vectors are selected to find the matching spectrogram.

TABLE I. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR FULL AND PARTIAL FEATURE VECTOR WHEN DCT IS APPLIED TO FULL IMAGE IN SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	54.16	59.16	56.66	56.66	68.33	62.50
192*192	58.33	65	67.5	65	73.33	69.16
128*128	65.83	64.16	71.66	67.5	74.16	72.5
64*64	70.83	70.83	71.66	72.50	77.50	75.83
32*32	75	73.33	74.16	75	80	77.5
20*20	78.33	75.33	78.33	71.66	81.66	80
16*16	72.5	76.66	74.16	74.16	76.66	79.16

Table II shows the identification rate for six sentences s1 to s6 when DCT is applied on full image in set B and partial feature vectors are selected to find the matching spectrogram.

TABLE II. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN DCT IS APPLIED TO FULL IMAGE IN SET B

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	73.33	70	76.67	75	78.33	78.33
128*128	78.33	73.33	80	78.33	81.67	81.67
64*64	80	80	78.33	86.67	83.33	88.33
32*32	90	86.67	86.67	86.67	86.67	90
20*20	86.67	86.67	86.67	88.33	90	90
16*16	85	85	86.67	86.67	91.67	90

Table III shows the comparison of overall identification rate considering all sentences, for partial feature vectors of different sizes when set A and set B is used. It also shows the number of DCT coefficients used for identifying speaker for corresponding selected portion of feature vector.

TABLE III. COMPARISON OF OVERALL IDENTIFICATION RATE FOR DIFFERENT NUMBER OF DCT COEFFICIENTS WHEN DCT IS APPLIED TO FULL IMAGE IN SET A AND SET B

Portion of feature vector selected	Number of DCT coefficients	% Identification rate	
		Set A	Set B
256*256	65536	60	70.83
192*192	36864	66.38	75.27
128*128	16384	69.30	78.88
64*64	4096	73.19	82.77
32*32	1024	75.83	87.77
20*20	400	77.63	88.05
16*16	256	76.66	87.5

2) Results for Walsh on full image

Results of Walsh transform on Spectrograms are tabulated below. Table IV shows the identification rate for sentences s1 to s6 for full and partial feature vectors when WALSH transform is applied on full image and set A is used.

TABLE IV. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN WALSH TRANSFORM IS APPLIED TO FULL IMAGE FROM SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	54.16	59.16	57.5	57.5	68.33	63.33
192*192	59.16	66.66	65.83	63.33	73.33	71.66
128*128	65.83	66.66	70.83	73.33	75	72.5
64*64	74.16	73.33	76.66	75	75.83	75
32*32	70.83	71.66	71.66	70.83	78.33	76.66
20*20	70.83	69.67	71.67	70.83	76.67	78.33
16*16	70	70.83	71.67	66.67	75	77.5

Table V shows sentencewise identification rate for WALSH transform on full image from set B. It can be observed from Table IV and Table V that, identification rate for each sentence is increased as more training is provided to the system. From both the tables, it can be seen that as size of the partial feature vector is decreased, the identification rate also decreases, achieves its peak value and then again decrease.

TABLE V. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR FULL AND PARTIAL FEATURE VECTOR WHEN WALSH TRANSFORM IS APPLIED TO FULL IMAGE FROM SET B

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	75	71.67	76.67	73.33	78.33	81.67
128*128	80	75	78.33	83.33	81.67	81.67
64*64	86.67	83.33	81.67	85	83.33	85
32*32	86.67	81.67	81.67	88.33	83.33	91.67
20*20	91.67	78.33	83.33	85	86.67	83.33
16*16	86.67	85	83.33	85	83.33	86.67

Table VI shows the overall identification rate considering all sentences, for partial feature vectors. For set A, highest identification rate is obtained for partial feature vector of size 64*64 i.e. 4096 WALSH coefficients. For set B, it requires 32*32 partial feature vector i.e. 1024 WALSH coefficients.

TABLE VI. COMPARISON OF OVERALL IDENTIFICATION RATE FOR VARYING NUMBER OF COEFFICIENTS WHEN WALSH TRANSFORM IS APPLIED TO FULL IMAGE FROM SET A AND SET B

Portion of feature vector selected	Number of Walsh coefficients	% Identification rate	
		Set A	Set B
256*256	65536	60	70.83
192*192	36864	66.66	76.11
128*128	16384	70.69	80
64*64	4096	75	84.16
32*32	1024	73.33	85.55
20*20	400	72.91	84.72
16*16	256	71.94	85

3) Results for HAAR on full image

Table VII shows sentencewise identification rate when 2-D HAAR transform is applied to full image with size 256*256 and partial feature vectors are selected from these feature vectors. These results are for set A.

TABLE VII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN HAAR TRANSFORM IS APPLIED TO FULL IMAGE FROM SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	54.16	59.16	57.5	57.5	68.33	63.33
192*192	59.16	66.66	65.83	63.33	73.33	71.66
128*128	65.83	66.66	70.83	73.33	75	72.5
64*64	74.16	73.33	76.66	75	75.83	75
32*32	70.83	71.66	71.66	70.83	78.33	76.66
20*20	65.83	73.33	71.67	70	75	76.67
16*16	70	70.83	71.67	66.67	75	77.5

Table VIII shows identification rate for HAAR transform on full image when set B is used. From both the tables, it can be seen that as the number of coefficients selected from the feature vector is decreased, the identification rate also decreases, achieves its peak value and then again decrease. From the Table VII and Table VIII, it can also be noted that when more training is provided to the system, identification rate per sentence is increased.

TABLE VIII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR WHEN HAAR TRANSFORM IS APPLIED TO FULL IMAGE WITH TRAINING SET OF EIGHT IMAGES FOR EACH SPEAKER

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
256*256	63.33	66.67	75	66.67	76.67	76.67
192*192	80	73.33	78.33	76.67	78.33	78.33
128*128	80	75	78.33	83.33	81.67	81.67
64*64	86.67	83.33	81.67	85	83.33	85
32*32	86.67	81.67	81.67	88.33	83.33	91.67
20*20	86.67	88.33	86.67	85	85	86.67
16*16	86.67	85	83.33	85	83.33	86.67

Table IX shows identification rate obtained by considering all six sentences, for set A and set B, with different sized partial feature vectors. Maximum identification rate is observed for 4096 and 400 HAAR coefficients with set A and set B respectively.

TABLE IX. COMPARISON OF OVERALL IDENTIFICATION RATE FOR VARYING NUMBER OF COEFFICIENTS WHEN HAAR TRANSFORM IS APPLIED TO FULL IMAGE FROM SET A AND SET B

Portion of feature vector selected	Number of HAAR coefficients	Identification rate (%)	
		Set A	Set B
256*256	65536	60	70.83
192*192	36864	67.91	77.5
128*128	16384	70.69	80
64*64	4096	75	84.16
32*32	1024	73.33	85.55
20*20	400	72.08	86.39
16*16	256	71.94	85

Table X shows the comparison of identification rates for all three transformation techniques on full image when set A and set B are used per speaker.

TABLE X. COMPARISON OF IDENTIFICATION RATES WHEN DCT, WALSH AND HAAR ON FULL IMAGE FROM SET A AND SET B

Portion of feature vector selected	Identification rate (%) when set A is used			Identification rate (%) when set B is used		
	DCT	WALSH	HAAR	DCT	WALSH	HAAR
256*256	60	60	60	70.83	70.83	70.83
192*192	66.38	66.66	67.91	75.27	76.11	77.5
128*128	69.30	70.69	70.69	78.88	80	80
64*64	73.19	75	75	82.77	84.16	84.16
32*32	75.83	73.33	73.33	87.77	85.55	85.55
20*20	77.63	72.91	72.08	88.05	84.72	86.39
16*16	76.66	71.94	71.94	87.5	85	85

B. Results for DCT/WALSH/HAAR on image block:

1) Results for DCT on image blocks:

Table XI shows the identification rate for sentences s1 to s6 when full and partial feature vectors are selected to identify speaker using DCT on image blocks using set A. Table XII shows the sentence wise identification rate for full and partial feature vectors when DCT is applied on image blocks for set B. It can be seen from the table that identification rate is improved when more training is provided to the system. From both the tables, it can be seen that, as the number of coefficients used for identification purpose decreases, the identification rate

increases, reaches its maximum value and then again decreases or remains constant.

TABLE XI. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR FULL AND PARTIAL FEATURE VECTOR USING DCT ON IMAGE BLOCKS FOR IMAGES FROM SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	54.16	59.16	57.5	57.5	68.33	63.33
96*384	60	63.33	65.33	65	73.33	68.33
64*256	65	65	70.83	66.66	74.16	71.16
32*128	70.83	70.83	70.83	71.66	76.66	75
16*64	75.83	74.16	75	75.83	81.66	77.5
8*32	69.16	76.66	75	75.83	75	75.83

TABLE XII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR FULL AND PARTIAL FEATURE VECTOR USING DCT ON IMAGE BLOCKS FOR IMAGES FROM SET B

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	71.67	70	76.67	75	78.33	78.33
64*256	78.33	73.33	80	76.67	81.67	81.67
32*128	78.33	80	78.33	86.67	83.33	86.67
16*64	90	88.33	86.67	90	86.67	88.33
8*32	88.33	88.33	85	86.67	90	86.67

Table XIII shows the comparison of overall identification rate considering all sentences, for partial feature vectors using DCT on image blocks. For both the training sets, maximum identification rate is achieved for partial feature of size 16*64 i.e. for 1024 DCT coefficients.

TABLE XIII. COMPARISON OF OVERALL IDENTIFICATION RATE FOR FULL AND PARTIAL FEATURE VECTOR PORTION USING DCT ON IMAGE BLOCKS FOR IMAGES FROM SET A AND SET B

Portion of feature vector selected	Number of DCT coefficients	Identification rate (%)	
		Set A	Set B
128*512	65536	60	70.83
96*384	36864	65.97	75
64*256	16384	68.88	78.61
32*128	4096	72.63	82.22
16*64	1024	76.66	88.33
8*32	256	74.58	86.67

2) Results for WALSH on image blocks:

Table IVX on the next page shows the sentencewise identification rate when WALSH transform is applied to image blocks using images in Set A.

TABLE IVX. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING WALSH ON IMAGE BLOCKS WITH IMAGES FROM SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	54.16	59.17	57.5	57.5	68.33	63.33
96*384	59.17	66.67	65.83	63.33	73.33	71.67
64*256	65.83	66.67	70.83	73.33	75	72.5
32*128	74.17	73.33	76.67	75	75.83	75
16*64	70.83	71.67	71.67	70.83	78.33	76.67
8*32	70	70.83	71.67	66.67	75	77.5

Table XV show the sentencewise identification rate when WALSH transform is applied to image blocks using Set B. From Table IVX and Table XV, it can be seen that, as the number of coefficients used for identification purpose decreases, the identification rate increases, reaches its maximum value and then again decreases or remains constant. Table XVI summarizes overall identification rate for both training sets for various partial feature vectors.

TABLE XV. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING WALSH ON IMAGE BLOCKS WITH IMAGES FROM SET B

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	75	71.67	76.67	73.33	78.33	81.67
64*256	80	75	78.33	83.33	81.67	81.67
32*128	86.67	83.33	81.67	85	83.33	85
16*64	86.67	81.67	81.67	88.33	83.33	91.67
8*32	86.67	85	83.33	85	83.33	86.67

TABLE XVI. COMPARISON OF OVERALL IDENTIFICATION RATE FOR VARYING SIZE OF FEATURE VECTOR PORTION USING WALSH ON IMAGE BLOCKS FOR IMAGES IN SET A AND SET B

Portion of feature vector selected	Number of WALSH coefficients	Identification rate (%)	
		Set A	Set B
128*512	65536	60	70.83
96*384	36864	66.67	76.11
64*256	16384	70.69	80
32*128	4096	75	84.16
16*64	1024	73.33	85.55
8*32	256	71.94	85

It can be observed from Table XVI that the maximum identification rate in case of Set A is obtained for 4096 WALSH coefficients i.e. for partial feature vector of size 32*128. The maximum identification rate in case of Set B is obtained for 1024 WALSH coefficients i.e. for partial feature vector of size 16*64.

3) Results for HAAR on image blocks:

Table XVII shows identification rate for each sentence when 2-D HAAR transform is applied on image blocks obtained by dividing an image into four equal and non-overlapping blocks as shown in Fig. 2. These results are for training Set A.

TABLE XVII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR VARYING PORTION OF FEATURE VECTOR USING HAAR ON IMAGE BLOCKS USING SET A

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	53.33	59.17	57.5	57.5	68.33	63.33
96*384	58.33	61.67	65.83	68.33	72.5	69.17
64*256	65.83	66.67	70.83	73.33	75	72.5
32*128	74.17	73.33	76.67	76.67	75	75.83
16*64	70.83	71.67	71.67	70.83	78.33	76.67
8*32	70	70.83	71.67	66.67	75	77.5

Table XVIII shows results when Set B is used and 2-D HAAR transform is applied on image blocks. For both the training sets, it is observed that, as the number of coefficients

used for identification purpose decreases, the identification rate increases, reaches some peak value and then again decreases or remains constant.

TABLE XVIII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR FULL AND PARTIAL FEATURE VECTOR USING HAAR ON IMAGE BLOCKS USING SET B

Portion of feature vector selected	Sentence					
	S1	S2	S3	S4	S5	S6
128*512	63.33	66.67	75	66.67	76.67	76.67
96*384	73.33	70	78.33	76.67	78.33	81.67
64*256	80	75	78.33	83.33	81.67	81.67
32*128	86.67	83.33	81.67	85	83.33	85
16*64	86.67	81.67	81.67	88.33	83.33	91.67
8*32	86.67	85	83.33	85	83.33	86.67

Table XIX shows overall identification rate for both training sets obtained by considering the identification rate for each sentence for various partial feature vectors. For Set A, the maximum identification rate of 75.27% is obtained for 32*128 feature vector. Whereas, for Set B, the maximum identification rate of 85.55% is obtained for 16*64 feature vector. Table XX shows comparison of overall identification rates for all three transformation techniques when applied on image blocks for Set A and Set B.

TABLE XIX. COMPARISON OF OVERALL IDENTIFICATION RATE FOR VARYING SIZE OF FEATURE VECTOR PORTION USING HAAR ON IMAGE BLOCKS USING SET A AND SET B

Portion of feature vector selected	Number of HAAR coefficients	Identification rate (%)	
		Set A	Set B
128*512	65536	59.86	70.83
96*384	36864	65.97	76.39
64*256	16384	70.69	80
32*128	4096	75.27	84.44
16*64	1024	73.33	85.55
8*32	256	71.94	85.27

TABLE XX. COMPARISON OF IDENTIFICATION RATES WHEN DCT, WALSH AND HAAR ARE APPLIED ON IMAGE BLOCKS FOR IMAGES IN SET A AND SET B

Portion of feature vector selected	Identification rate (%) When Set A is used			Identification rate (%) When Set B is used		
	DCT	WALSH	HAAR	DCT	WALSH	HAAR
128*512	60	60	59.86	70.83	70.83	70.83
96*384	65.97	66.67	65.97	75	76.11	76.39
64*256	68.88	70.69	70.69	78.61	80	80
32*128	72.63	75	75.27	82.22	84.16	84.44
16*64	76.66	73.33	73.33	88.33	85.55	85.55
8*32	74.58	71.94	71.94	86.67	85	85.27

C. Results for DCT/ WALSH/ HAAR on Row Mean of an image and Row Mean of image blocks :

1) Results for DCT on Row Mean of an image :

Table XXI shows sentence wise results obtained for Set A when DCT of Row Mean is taken by dividing an image into different number of non-overlapping blocks.

TABLE XXI. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR DCT ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS USING SET A

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	57.5	66.66	64.16	60.83	60.83	62.5
4 Blocks (128*128)	60.83	70.83	63.33	65.83	70	65.83
16 Blocks (64*64)	69.16	75.83	70.83	65.83	73.33	71.66
64 Blocks (32*32)	75	76.66	75.83	70	78.83	75.83
256 Blocks (16*16)	76.66	75	75.83	72.5	80	82.5
1024 Blocks (8*8)	74.16	72.5	75	72.5	80.83	78.33

It can be seen from the Table XXI that, as the block size chosen for calculating Row Mean reduces, better identification rate is achieved. For block size 16*16, maximum identification rate is obtained and then it decreases again.

Table XXII shows the identification rate for sentence s1 to s6 and Set B of images. It can be seen from the Table XXII that, as the block size chosen for calculating Row Mean reduces, better identification rate is achieved. For block size 16*16, maximum identification rate is obtained and then it decreases again.

TABLE XXII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR DCT ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS USING SET B

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	73.33	76.67	78.33	76.67	75	80
4 Blocks (128*128)	80	80	78.33	81.67	81.67	80
16 Blocks (64*64)	91.67	81.67	83.33	83.33	81.67	83.33
64 Blocks (32*32)	91.67	85	86.67	86.67	86.67	88.33
256 Blocks (16*16)	91.67	88.33	88.33	85	91.67	90
1024 Blocks (8*8)	88.33	83.33	85	86.67	85	88.33

The overall identification rates for both the sets, when DCT of Row Mean is taken by dividing an image into different number of non-overlapping blocks are tabulated in Table XXIII.

TABLE XXIII. COMPARISON OF OVERALL IDENTIFICATION RATE FOR DCT ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS WITH SET A AND SET B

No. of blocks for image split	Number of DCT coefficients	Identification rate (%)	
		For Set A	For Set B
Full image (256*256)	256	62.08	76.67
4 Blocks 128*128)	512	66.11	80.27
16 Blocks (64*64)	1024	71.11	84.17
64 Blocks (32*32)	2048	75.27	87.5
256 Blocks 16*16)	4096	77.08	89.17
1024 Blocks (8*8)	8192	75.55	86.11

2) Results for WALS on Row Mean of an image :

Table IVXX shows the sentence wise identification rate when Walsh transform is applied to Row Mean of an image

when it is divided into different number of non-overlapping and Set A is used. Table XXV shows the sentence wise identification rate when Walsh transform is applied to Row Mean of an image when it is divided into different number of non-overlapping and Set B.

TABLE IVXX. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR WALSH TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS WITH SET A

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	57.5	66.66	64.16	60.83	60.83	62.5
4 Blocks (128*128)	60.83	70.83	63.33	65.83	70	65.83
16 Blocks (64*64)	69.16	75.83	70.83	65.83	73.33	71.66
64 Blocks (32*32)	75	76.66	75.83	70	78.83	75.83
256 Blocks (16*16)	76.66	75	75.83	72.5	80	82.5
1024 Blocks (8*8)	74.16	72.5	75	72.5	80.83	78.33

TABLE XXV. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR WALSH TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS WITH SET B

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	73.33	76.66	78.33	76.66	75	80
4 Blocks (128*128)	80	80	78.33	81.67	81.67	80
16 Blocks (64*64)	91.67	81.67	83.33	83.33	81.67	83.33
64 Blocks (32*32)	91.67	85	86.66	86.66	86.66	88.33
256 Blocks (16*16)	91.67	88.33	88.33	85	91.67	90
1024 Blocks (8*8)	88.33	83.33	85	86.66	85	88.33

It can be seen from the Table IVXX and Table XXV that, as the block size chosen for calculating Row Mean reduces, better identification rate is achieved.

Table XXVI summarizes overall identification rate for both training sets by considering all six sentences. For block size 16*16, maximum identification rate is obtained and then it decreases again.

TABLE XXVI. COMPARISON OF OVERALL IDENTIFICATION RATE FOR WALSH TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS FOR SET A AND SET B

No. of blocks for image split	Number of Walsh coefficients	Identification rate (%)	
		Set A	Set B
Full image (256*256)	256	62.08	76.67
4 Blocks (128*128)	512	66.11	80.27
16 Blocks (64*64)	1024	71.11	84.17
64 Blocks (32*32)	2048	75.27	87.5
256 Blocks (16*16)	4096	77.08	89.17
1024 Blocks (8*8)	8192	75.55	86.11

3) Results for HAAR on Row Mean of an image :

Table XXVII shows identification rate for each sentence when 1-D HAAR transform is applied to Row Mean of an 256*256 image and when image is divided into different number of non-overlapping blocks for Set A. Similarly, Table XXVIII shows identification rate for each sentence when 1-D HAAR transform is applied to Row Mean of an 256*256 image

and when image is divided into different number of non-overlapping blocks for Set B.

TABLE XXVII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR HAAR TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS USING SET A

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	57.5	66.66	64.16	60.83	60.83	62.5
4 Blocks (128*128)	60.83	70.83	63.33	65.83	70	65.83
16 Blocks (64*64)	69.16	75.83	70.83	65.83	73.33	71.66
64 Blocks (32*32)	75	76.66	75.83	70	78.83	75.83
256 Blocks (16*16)	76.66	75	75.83	72.5	80	82.5
1024 Blocks (8*8)	74.16	72.5	75	72.5	80.83	78.33

TABLE XXVIII. IDENTIFICATION RATE FOR SENTENCES S1 TO S6 FOR HAAR TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS USING SET B

No. of blocks for image split	Sentence					
	S1	S2	S3	S4	S5	S6
Full image (256*256)	73.33	76.67	78.33	76.67	75	80
4 Blocks (128*128)	80	80	78.33	81.67	81.67	80
16 Blocks (64*64)	91.67	81.67	83.33	83.33	81.67	83.33
64 Blocks (32*32)	91.67	85	86.67	86.67	86.67	88.33
256 Blocks (16*16)	91.67	88.33	88.33	85	91.67	90
1024 Blocks (8*8)	88.33	83.33	85	86.67	85	88.33

Table XXIX shows overall identification rate for the two training sets when 1-D HAAR transform is applied to an image divided into different number of equal and non-overlapping blocks.

TABLE XXIX. COMPARISON OF OVERALL IDENTIFICATION RATE FOR HAAR TRANSFORM ON ROW MEAN OF AN IMAGE WHEN IMAGE IS DIVIDED INTO DIFFERENT NUMBER OF NON-OVERLAPPING BLOCKS FOR SET A AND B.

No. of blocks for image split	Number of HAAR coefficients	Identification rate (%)	
		For Set A	For Set B
Full image (256*256)	256	62.08	76.67
4 Blocks (128*128)	512	66.11	80.27
16 Blocks (64*64)	1024	71.11	84.17
64 Blocks (32*32)	2048	75.27	87.5
256 Blocks (16*16)	4096	77.08	89.17
1024 Blocks (8*8)	8192	75.55	86.11

Overall identification rate for DCT, WALSH and HAAR on Row Mean of an image and image blocks are summarized in the Table XXX.

TABLE XXX. COMPARISON OF DCT, WALSH AND HAAR ON ROW MEAN OF IMAGE AND IMAGE BLOCKS

No. of blocks for image split	Identification rate (%) when Set A is used			Identification rate (%) when Set B is used		
	DCT	WALSH	HAAR	DCT	WALSH	HAAR
Full image (256*256)	62.08	62.08	62.08	76.67	76.67	76.67
4 Blocks (128*128)	66.11	66.11	66.11	80.27	80.27	80.27
16 Blocks (64*64)	71.11	71.11	71.11	84.17	84.17	84.17
64 Blocks (32*32)	75.27	75.27	75.27	87.5	87.5	87.5
256 Blocks (16*16)	77.08	77.08	77.08	89.17	89.17	89.17
1024 Blocks (8*8)	75.55	75.55	75.55	86.11	86.11	86.11

V. COMPLEXITY ANALYSIS

A. Complexity analysis of DCT, WALSH and HAAR on full image:

For 2-D DCT on $N \times N$ image, $2N^3$ multiplications are required and $2N^2(N-1)$ additions are required. For 2-D WALSH on $N \times N$ image, $2N^2(N-1)$ additions are required. For 2-D HAAR transform on $N \times N$ image where $N=2m$, number of multiplications required are $2(m+1)N^2$ and number of additions required are $2mN^2$. Table XXXI summarizes these details along with actual values of mathematical computations needed for processing of 256*256 images.

TABLE XXXI. COMPARISON BETWEEN DCT, WALSH AND HAAR WITH RESPECT TO MATHEMATICAL COMPUTATIONS AND IDENTIFICATION RATE WHEN APPLIED ON FULL IMAGE

Parameter	Algorithm		
	DCT on full image($N \times N$)	WALSH on full image($N \times N$)	HAAR on full image ($N \times N$)
Number of Multiplications	$2N^3$	0	$2(m+1)N^2$
N=256	33554432	0	1179648
Number of Additions	$2N^2(N-1)$	$2N^2(N-1)$	$2mN^2$
N=256	33423360	33423360	1048576
Identification rate (%) for Set A	77.63	75	75
Identification rate (%) for Set B	88.05	85.55	86.39

From the above table, it can be seen that DCT on full image gives the highest identification rate for both the training sets as compared to WALSH and HAAR on full image. However this outstanding performance is achieved at the expense of higher computations.

Number of multiplications required by DCT on full image is approximately 28 times more than the number of multiplications required by HAAR on full image. Whereas number of additions required by DCT on full image is approximately 31 times more than the number of additions required by HAAR on full image.

Though WALSH on full image does not require any multiplications, overall CPU time taken by it is more than that of HAAR on full image. This is because the number of additions taken by WALSH on full image is approximately 31 times more than the number of additions required by HAAR on full image.

B. Complexity analysis of DCT, WALSH and HAAR on image blocks:

The number of multiplications required in case of 2-D DCT on image blocks is N^3 and the number of additions required are $N^2(N-2)$.

For 2-D WALSH on four image blocks of size $N/2 \times N/2$, number of additions required are $N^2(N-2)$.

The number of multiplications required for 2-D HAAR on image blocks is $2mN^2$. Similarly number of additions required for 2-D HAAR on image blocks is $2(m-1)N^2$. Table XXXII summarizes these details along with actual values of mathematical computations needed for processing of 256×256 images.

TABLE XXXII COMPARISON BETWEEN DCT, WALSH AND HAAR WITH RESPECT TO MATHEMATICAL COMPUTATIONS AND IDENTIFICATION RATE WHEN APPLIED ON IMAGE BLOCKS

Parameter	Algorithm		
	DCT on image blocks ($N/2 \times N/2$)	Walsh on image blocks ($N/2 \times N/2$)	HAAR on image blocks ($N/2 \times N/2$)
Number of Multiplications	N^3	0	$2(m+1)N^2$
N=256, four blocks	16777216	0	1048576
Number of Additions	$N^2(N-2)$	$N^2(N-2)$	$2(m-1)N^2$
N=256, four blocks	16646144	16646144	917504
Identification rate (%) for Set A	76.66	75	75.27
Identification rate (%) for Set B	88.33	85.55	85.55

From the Table XXXII, it can be seen that, in all the three transformation techniques on image blocks, DCT on image blocks gives best identification rate for both the training sets. But this performance is achieved at the expense of higher number of computations. DCT on image blocks takes 16 times more multiplications and 18 times more additions than HAAR on image blocks. Though WALSH transform does not need any multiplications, still it takes more number of computations than HAAR. This is because WALSH on image blocks requires approximately 18 times more additions than HAAR on image blocks.

C. Complexity Analysis of DCT, Walsh and HAAR on Row Mean of an image and on Row Mean of image blocks:

Since Row Mean of an image is a one dimensional vector, only 1-D DCT, WALSH and HAAR need to be applied on Row Mean. This itself reduces the number of multiplications and additions required for feature vector calculation. Row Mean of an image of size $N \times N$ is a vector of size $N \times 1$. For 1-D DCT on this $N \times 1$ vector, N^2 multiplications and $N(N-1)$

additions are required. One dimensional Walsh on Row Mean of an image takes $N(N-1)$ additions and no multiplications. Whereas, 1-D HAAR on Row Mean of an image of size $N \times N$ requires $(m+1)N$ multiplications and mN additions where $N=2^m$. Following Table XXXIII summarizes this statistics in case of each transformation technique applied for the Row Mean of block size 16×16 which gives highest identification rate.

TABLE XXXIII. COMPARISON BETWEEN DCT, WALSH AND HAAR WITH RESPECT TO MATHEMATICAL COMPUTATIONS AND IDENTIFICATION RATE WHEN APPLIED ON ROW MEAN OF AN IMAGE

Parameter	Algorithm		
	DCT on Row Mean of image ($N \times 1$)	Walsh on Row Mean of image ($N \times 1$)	HAAR on Row Mean of image ($N \times 1$)
Number of Multiplications	N^2	0	$(m+1)N$
N=16, 256 blocks	65536	0	20480
Number of Additions	$N(N-1)$	$N(N-1)$	mN
N=16, 256 blocks	61440	61440	16384
Identification rate (%) for Set A	77.08	77.08	77.08
Identification rate (%) for Set B	89.17	89.17	89.17

From the Table XXXIII, we can see that all three transformation techniques result in same identification rate when applied on the Row Mean of an image and on Row Mean of an image blocks. For both the training sets, highest identification rate is obtained when image is divided into 16×16 size blocks. However, in terms of computations, HAAR transform is proved to be better one. Number of multiplications required by HAAR is approximately three times less than the number of multiplications required in case of DCT on Row Mean. Also the number of additions required by HAAR is 3.5 times less than the number of additions required by DCT and WALSH on Row Mean of an image.

Along with the different approaches of applying transformation techniques on spectrograms, comparative study of computational complexity of three transformation techniques for each approach has been done and is presented below.

D. Complexity Analysis of DCT transform on Full, Block and Row Mean of Spectrograms:

For 2-D DCT on $N \times N$ image, $2N^3$ multiplications are required and $2N^2(N-1)$ additions are required. For 2-D DCT on four blocks of size $N/2 \times N/2$, N^3 multiplications are required and $N^2(N-2)$ additions are required. For 1-D DCT on $N \times 1$ image, N^2 multiplications are needed and $N(N-1)$ additions are needed. These computational details are summarized in Table XXXIV along with the actual number of computations for 256×256 image using three methods of applying DCT.

TABLE XXXIV. COMPUTATIONAL DETAILS FOR 2-D DCT ON N*N IMAGE, 2-D DCT ON N/2*N/2 IMAGE AND 1-DCT ON N*1 IMAGE RESPECTIVELY

Parameter → Algorithm ↓	No. of Multiplications	No. of Additions
2-D DCT on N*N image	$2N^3$	$2N^2(N-1)$
2-D DCT on 256*256 image	33554832	33424159
2-D DCT on four blocks of size N/2*N/2 each	N^3	$N^2(N-2)$
2-D DCT on four blocks of size 256/2*256/2 each	16778240	16648191
1-D DCT on N*1 Row Mean of N*N image	N^2	$N(N-1)$
1-D DCT on N*1 Row Mean of 256*1 image	69632	69631

When all three methods of applying DCT are compared, it has been observed that though number of coefficients used in Row Mean method is higher, number of multiplications and additions reduce drastically as compared to other two methods. Number of multiplications in DCT on full image method is 480 times more than the number of multiplications in Row Mean method whereas for DCT on image blocks it is 241 times more. Number of additions needed in DCT on full image and DCT on image blocks are also 480 times and 239 times more than the additions required in Row mean method respectively. For the Set A, the identification rate for DCT on Row Mean is almost same as identification rate for DCT on full image. In case of Set B, DCT on Row Mean gives better identification rate as compared to DCT on full image and DCT on image blocks and that too with reduced number of mathematical computations.

E. Complexity Analysis of WALSH transform on Full, Block and Row Mean of Spectrograms:

For 2-D WALSH on N*N image, $2N^2(N-1)$ additions are required. For 2-D WALSH on four blocks of size N/2*N/2, $N^2(N-1)$ additions are required. Whereas for 1-D WALSH on N*1 image, $N(N-1)$ additions are needed as shown in table XXXV. In all three cases number of multiplications required is zero.

TABLE XXXV. COMPUTATIONAL DETAILS FOR 2-D WALSH ON N*N IMAGE, 2-D WALSH ON N/2*N/2 IMAGE AND 1-D WALSH ON N*1 IMAGE RESPECTIVELY

Parameter → Algorithm ↓	No. of Multiplications	No. of Additions
2-D WALSH on N*N image	0	$2N^2(N-1)$
2-D WALSH on 256*256 image	0	33423360
2-D WALSH on four blocks of size N/2*N/2 each	0	$N^2(N-2)$
2-D WALSH on four blocks of size 256/2*256/2 each	0	16646144
1-D WALSH on N*1 size Row Mean vector of image N*N	0	$N(N-1)$
1-D WALSH on 256*1 size Row Mean vector of image 256*1	0	65280

From table 7.5 it can be seen that number of additions required when WALSH transform is applied on full image is 512 times more than the number of additions required when

WALSH transform is applied to Row Mean of an image. Also the number of additions required when WALSH transform is applied on image blocks is 255 times more than the number of additions required when WALSH transform is applied to Row Mean of an image. Thus number of additions is drastically reduced for Walsh transform on Row Mean of an image.

F. Complexity Analysis of HAAR transform on Full, Block and Row Mean of Spectrograms:

For 2-D HAAR transform on N*N image where $N=2^m$, number of multiplications required are $2(m+1)N^2$ and number of additions required are $2mN^2$. For 2-D HAAR transform on four blocks of size N/2*N/2 each, $2mN^2$ multiplications and $2(m-1)N^2$ additions are needed. Whereas for 1-D HAAR transform on N*1 image, number of multiplications required are $(m+1)N$ and number of additions are mN as shown in table XXXVI.

TABLE XXXVI. COMPUTATIONAL DETAILS FOR 2-D HAAR ON N*N IMAGE, 2-D HAAR ON N/2*N/2 IMAGE AND 1-D HAAR ON N*1 IMAGE RESPECTIVELY

Parameter → Algorithm ↓	No. of Multiplications	No. of Additions
2-D HAAR on N*N image	$2(m+1)N^2$	$2mN^2$
2-D HAAR on 256*256 image	1179648	1048576
2-D HAAR on four blocks of size N/2*N/2 each	$2mN^2$	$2(m-1)N^2$
2-D HAAR on four blocks of size 256/2*256/2 each	1048576	917504
1-D HAAR on N*1 image	$(m+1)N$	mN
1-D HAAR on 256*1 size Row Mean vector of image 256*1	2304	2048

From Table XXXVI, it can be seen that number of multiplications required when HAAR transform is applied on full image is 512 times more than the number of multiplications required when HAAR transform is applied to Row Mean of an image. Also the number of multiplications required when HAAR transform is applied on image blocks is 455 times more than the number of multiplications required when HAAR transform is applied to Row Mean of an image. Thus number of multiplications is drastically reduced for HAAR transform on Row Mean of an image. Number of additions required is also reduced to a greater extent when transformation technique is applied on Row Mean of an image. Number of additions required when HAAR transform is applied on full image is 512 times more than the number of additions required when HAAR transform is applied to Row Mean of an image. Also the number of additions required when HAAR transform is applied on image blocks is 448 times more than the number of additions required when HAAR transform is applied to Row Mean of an image.

VI. CONCLUSION

In this paper, closed set text dependent speaker identification has been considered using three different transformation techniques namely DCT, WALSH and HAAR. Each transformation technique is applied in three ways:

- On full image

- b) On image blocks and
- c) On Row Mean of an image.

For each method, two training sets were used as mentioned earlier.

It can be clearly concluded from the results that as more training is provided to the system, more accuracy is obtained in the results in terms of identification rate.

Further for each method, Identification rates are obtained for various numbers of coefficients from feature vectors of images. It has been observed that as the number of coefficients chosen is smaller up to a certain limit; better identification rate is achieved in all three methods.

DCT on full image gives its best identification rate for only 20*20 portion of feature vector i.e. by using only 400 DCT coefficients. DCT on image blocks gives highest identification rate when 16*64 portion of its feature vector is considered which has 1024 DCT coefficients. Finally DCT on Row Mean gives highest identification rate for Row Mean of 16*16 size image blocks i.e. for 4096 DCT coefficients. When these highest identification rates in all three methods in DCT are compared, it has been observed that DCT on image blocks gives slightly improved results for training set of eight images per speaker. Whereas, DCT on Row Mean, further improves these results with drastically reduced computations. Though the number of coefficients used in Row Mean method is higher, overhead caused for its comparison is negligible as compared to number of mathematical computations needed in other two approaches.

Similarly, WALSH on Row Mean of image gives better identification rates as compared to WALSH on full image and WALSH on image blocks for both the training sets. These better identification rates are obtained with the advantage of reduced mathematical computations. For HAAR transform also, identification rate for HAAR on Row Mean is better than HAAR on full image and HAAR on image blocks.

From the results of DCT, WALSH and HAAR on full image, it can be concluded that DCT on full image gives better identification rate than WALSH and HAAR on full image but at the expense of large number of mathematical computations. In WALSH transform on full image, numbers of mathematical computations required are greatly reduced as compared to DCT since no multiplications are required in WALSH. These computations are further reduced by use of HAAR transform but at the slight expense of identification rate. Similar conclusions can be drawn for DCT, WALSH and HAAR on image blocks. So there is a trade off between better identification rate and less CPU time for mathematical computations.

However, in case of Row Mean approach of applying transform, performances of all three transformation techniques are same for a specific block size chosen for Row Mean. In that HAAR transform proves to be better because it requires minimum number of computations.

The overall conclusion is that Row Mean technique requires less number of mathematical computations and hence less CPU time for all three transformation techniques as

compared to transformation techniques on full image and on image blocks. HAAR transform on Row Mean of an image gives the best result with respect to identification rate as well as number of computations required.

REFERENCES

- [1] Evgeniy Gabrilovich, Alberto D. Berstin: "Speaker recognition: using a vector quantization approach for robust text-independent speaker identification", Technical report DSPG-95-9-001', September 1995.
- [2] Tridibesh Dutta, "Text dependent speaker identification based on spectrograms", Proceedings of Image and vision computing, pp. 238-243, New Zealand 2007..
- [3] J.P.Campbell, "Speaker recognition: a tutorial", Proc. IEEE, vol. 85, no. 9, pp. 1437-1462, 1997.
- [4] D. O'Shaughnessy, "Speech communications- Man and Machine", New York, IEEE Press, 2nd Ed., pp. 199, pp. 437-458, 2000.
- [5] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Transaction Acoustics Speech and Signal Processing, vol. 4, pp. 375-366, 1980.
- [6] Wang Yutai, Li Bo, Jiang Xiaoqing, Liu Feng, Wang Lihao, "Speaker Recognition Based on Dynamic MFCC Parameters", International Conference on Image Analysis and Signal Processing, pp. 406-409, 2009
- [7] Azzam Sleit, Sami Serhan, and Loai Nemir, "A histogram based speaker identification technique", International Conference on ICADIWT, pp. 384-388, May 2008.
- [8] B. S. Atal, "Automatic Recognition of speakers from their voices", Proc. IEEE, vol. 64, pp. 460-475, 1976.
- [9] Jialong He, Li Liu, and G'unther Palm, "A discriminative training algorithm for VQ-based speaker Identification", IEEE Transactions on speech and audio processing, vol. 7, No. 3, pp. 353-356, May 1999.
- [10] Debadatta Pati, S. R. Mahadeva Prasanna, "Non-Parametric Vector Quantization of Excitation Source Information for Speaker Recognition", IEEE Region 10 Conference, pp. 1-4, Nov. 2008.
- [11] Tridibesh Dutta and Gopal K. Basak, "Text dependent speaker identification using similar patterns in spectrograms", PRIP'2007 Proceedings, Volume 1, pp. 87-92, Minsk, 2007.
- [12] Andrew B. Watson, "Image compression using the Discrete Cosine Transform", Mathematica journal, 4(1), pp. 81-88, 1994.,
- [13] <http://www.itee.uq.edu.au/~conrad/vidtimit/>
- [14] <http://www2.imm.dtu.dk/~lf/elsdsr/>
- [15] H.B.Kekre, Sudeep D. Thepade, "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image", International Journal of Information Retrieval (IJIR), Serials Publications, Volume 2, Issue 1, pp. 72-79 (ISSN: 0974-6285), 2009.
- [16] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "DCT Applied to Row Mean and Column Vectors in Fingerprint Identification", In Proceedings of International Conference on Computer Networks and Security (ICCNS), 27-28 Sept. 2008, VIT, Pune.
- [17] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS), Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.
- [18] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.
- [19] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi, "Improved Texture Feature Based Image Retrieval using Kekre's Fast Codebook Generation Algorithm", Springer-International Conference on Contours of Computing Technology (Thinkquest-2010),

Babasaheb Gawde Institute of Technology, Mumbai, 13-14 March 2010, The paper will be uploaded on online Springerlink.

- [20] H. B. Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval by Kekre's Transform Applied on Each Row of Walsh Transformed VQ Codebook", (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, The paper is invited at ICWET 2010. Also will be uploaded on online ACM Portal.
- [21] H. B. Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 55-65. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).
- [22] H. B. Kekre, Ms. Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), Volume 9, Issue 5, pp.: 1-8, September 2009. Available online at <http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html>.
- [23] H. B. Kekre, Sudeep Thepade, Akshay Maloo, "Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform", International Journal of Engineering Science and Technology, Vol.. 2, No. 4, 2010, 362-371
- [24] H. B. Kekre, Sudeep Thepade, Akshay Maloo, "Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre's Transform", CSC-International Journal of Image processing (IJIP), Vol.. 4, No.2, pp.:142-155, May 2010.
- [25] H. B. Kekre, Tanuja Sarode "Two Level Vector Quantization Method for Codebook Generation using Kekre's Proportionate Error Algorithm", CSC-International Journal of Image Processing, Vol.4, Issue 1, pp.1-10, January-February 2010
- [26] H. B. Kekre, Sudeep Thepade, Akshay Maloo, "Eigenvectors of Covariance Matrix using Row Mean and Column Mean Sequences for Face Recognition", CSC-International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2), pp. 42-50, May 2010.
- [27] H. B. Kekre, Tanuja Sarode, Shachi Natu, Prachi Natu, "Performance Comparison Of 2-D DCT On Full/Block Spectrogram And 1-D DCT On Row Mean Of Spectrogram For Speaker Identification", (Selected), CSC-International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (3), pp. 100-112, August 2010, Malaysia..
- [28] H. B. Kekre, Tanuja Sarode, Shachi Natu, Prachi Natu, "Performance Comparison of Speaker Identification Using DCT, Walsh, Haar On Full And Row Mean Of Spectrogram", (Selected), International Journal of Computer Applications, pp. 30-37, August 2010, USA.
- [29] H. B. Kekre, Tanuja Sarode, Shachi Natu, Prachi Natu, "Speaker Identification using 2-d DCT, Walsh and Haar on Full and Block Spectrograms", International Journal of Computer Science and Engineering, Volume 2, Issue 5, pp. 1733-1740, August 2010.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle (w), Mumbai, INDIA. He ha guided 17 Ph.D.s, 150



M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 250 papers in National / International Conferences / Journals to his credit. Recently six students working under his guidance have received best paper awards. Currently he is guiding ten Ph.D. students.

Dr. Tanuja K. Sarode has received M.E. (Computer Engineering) degree from Mumbai University in 2004 and Ph.D. from Mukesh Patel School of Technology, Management and Engg. SVKM's NMIMS University, Vile-Parle (W), Mumbai, INDIA, in 2010. She has more than 10 years of experience in teaching. Currently working as Assistant Professor in Dept. of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT). Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has 70 papers in National /International Conferences/journal to her credit.



Shachi Natu has received B.E. (Computer) degree from Mumbai University with first class in 2004. Currently Pursing M.E. in Computer Engineering from University of Mumbai. She has 05 years of experience in teaching. Currently working as Lecturer in department of Information Technology at Thadomal Shahani Engineering College, Bandra (w), Mumbai. Her areas of interest are Image Processing, Data Structure, Database Management Systems and operating systems. She has 3 papers in National / International Conferences /journal to her credit.



Prachi Natu has received B.E. (Electronics and Telecommunication) degree from Mumbai University with first class in 2004. Currently Pursing M.E. in Computer Engineering from University of Mumbai. She has 04 years of experience in teaching. Currently working as Lecturer in Computer Engineering department at G. V. Acharya Institute of Engineering and Technology, Shelu. Mumbai. Her areas of interest are Image Processing, Database Management Systems and operating systems. She has 3 papers in National / International Conferences /journal to her credit.



A Research Proposal for Mitigating DoS Attacks in IP-based Networks

Sakharam Lokhande

Assistant Professor

School of Computational Science,
Swami Ramanand Teerth Marathwada University, Nanded,
MS, India, 431606.

Parag Bhalchandra

Assistant Professor

School of Computational Science,
Swami Ramanand Teerth Marathwada University, Nanded,
MS, India, 431606.

Nilesh Deshmukh

Assistant Professor

School of Computational Science,
Swami Ramanand Teerth Marathwada University, Nanded,
MS, India, 431606.

Dr. Santosh Khamitkar

Associate Professor

School of Computational Science,
Swami Ramanand Teerth Marathwada University, Nanded,
MS, India, 431606.

Santosh Phulari

Assistant Professor

School of Computational Science, Swami Ramanand Teerth
Marathwada University, Nanded,
MS, India, 431606.

Ravindra Rathod

Assistant Professor

School of Computational Science, Swami Ramanand Teerth
Marathwada University, Nanded,
MS, India, 431606.

Abstract : This paper studies denial of service (DoS) attacks in computer networks. These attacks are known for preventing availability of network services from their legitimate users. After careful review of literature, we wish to presents a structured view on possible attack and defense mechanisms. An outline to describe some new defense mechanisms is also presented in terms of a research proposal .

Keywords- Denial of Service Attacks, Intrusion, Security

I. PROBLEM DEFINATION

Defending against DoS attacks is a task from network and computer security. As scientific disciplines, network and computer security are relatively primitive. An indication of this fact is to be aware that the computer security terminology is not yet stabilized [4]. Computer and network security aspects were first studied in the early 1970s. As in some of the earliest security papers listed and available in, the Denial of Service attacks are timely and extremely important research topic. According to the CSI/FBI computer crime and security survey in the United States [1] for the year 2004, DoS attacks are the second most widely detected outsider attack type in computer networks, immediately after virus infections. A computer crime and security survey in Australia[1] for the year 2004, gives similar results. It is currently not possible to prevent DoS attacks because many of these attacks are based on using ordinary protocols and services in an overwhelming

manner. Specific security holes in the victim hosts or networks are thus not necessarily needed. For this reason we can only mitigate these attacks.

II. OVERVIEW OF DENIAL OF SERVICE ATTACKS

Denials of Service (DoS) attacks have proved to be a serious and permanent threat to users, organizations, and infrastructures of the Internet [1]. The primary goal of these attacks is to prevent access to a particular resource like a web server [2]. A large number of defenses against DoS attacks have been proposed in the literature, but none of them gives reliable protection. There will always be vulnerable hosts in the Internet to be used as sources of attack traffic. It is simply not feasible to expect all existing hosts in the Internet to be protected well enough. In addition, it is very difficult to reliably recognize and filter only attack traffic without causing any collateral damage to legitimate traffic.

A DoS attack can be carried out either as a flooding or a logic attack. A *Flooding DoS attack* is based on brute force. Real-looking but unnecessary data is sent as much as possible to a victim. As a result, network bandwidth is wasted, disk space is filled with unnecessary data (such as spam e-mail, junk files, and intentional error messages), fixed size data structures inside host software are filled with bogus information, or processing power is spent for un useful purposes. To amplify the effects, DoS attacks can be run in a coordinated fashion from several sources at the same time

(Distributed DoS, DDoS). A *logic DoS attack* is based on an intelligent exploitation of vulnerabilities in the target. For example, a skillfully constructed fragmented Internet Protocol (IP) datagram may crash a system due to a serious fault in the operating system (OS) software. Another example of a logic attack is to exploit missing authentication requirements by injecting bogus routing information to prevent traffic from reaching a victim's network. [5, 6]

There are two major reasons that make DoS attacks attractive for attackers. The first reason is that there are effective automatic tools available for attacking any victim, so expertise is not necessarily required. The second reason is that it is usually impossible to locate an attacker without extensive human interaction or without new features in most routers of the Internet. DoS attacks make use of vulnerabilities in end-hosts, routers, and other systems connected to a computer network. The size of a population having the same vulnerability can be large. In July 2003 a vulnerability was found from the whole population of Cisco routers and switches running any version of the Cisco IOS software and configured to process Internet Protocol version 4 (IPv4) packets. This vulnerability made it possible to block an interface, which resulted in a DoS condition without any alarms being triggered. Another example of a large population is the Microsoft Windows Metafile (WMF) vulnerability which was found in December 2005 from all versions of Windows 98, 98SE, ME, 2000, and XP. This vulnerability made it possible to install any malicious software on these hosts, for example, to send DoS attack traffic. User interaction was, however, required to exploit this vulnerability.

III. RESEARCH PROBLEM

Mitigating DoS attacks is difficult especially due to the following problems:

- 1) Very little has been done to compare, contrast, and categorize the different ideas related to DoS attacks and defenses. As a result it is difficult to understand what a computer network user needs to do and why to mitigate the threat from DoS attacks.
- 2) There are no effective defense mechanisms against many important DoS attack types. There is no guidance on how to select defense mechanisms.
- 3) Existing defense mechanisms have been evaluated according to very limited criteria.
- 4) Often relevant risks have been ignored (such as in [3]) or evaluations have been carried out under ideal conditions.
- 5) No research publications exist for giving a systematic list of issues related to defense evaluation

IV. OBJECTIVE OF THE RESEARCH

The objective of this research proposal is to help any user in any network for mitigating DoS attacks in IP-based networks. This study concentrates especially on the following areas:

- 1) One should understand existing attack mechanisms and available defense mechanisms, and have a rough idea about the benefits (best-case performance) of each defense mechanism.

- 2) One should acknowledge possible situation dependency of defense mechanisms, and be able to choose the most suitable defense when more than one defense mechanisms are available against a specific attack type.

- 3) One should evaluate defense mechanisms in a comprehensive way, including both benefits and disadvantages (worst-case performance), as an attacker can exploit any weakness in a defense mechanism.

Knowledge of all of these issues is necessary in successful mitigation of DoS attacks. Without knowing how a specific defense mechanism works under different possible conditions and what the real benefits and weaknesses are, it is not possible to assure the suitability of a defense mechanism against a certain type of a DoS attack.

V. RESEARCH METHODOLOGY

Research methodologies aimed to be used in this proposal, are primarily based on simulating different attack scenarios, but measurements, mathematical modeling based on game theory, and requirement specification are also planned to be used.

VI. SCOPE OF THE RESEARCH

Since this proposal studies DoS attacks in computer networks using the Internet Protocol (IP), namely the Internet and mobile ad hoc networks, is extremely useful for the security concern. DoS attacks in the physical world will not be studied here. Major work concentrate on the fixed (wired) Internet, but most of the considered attack and defense mechanisms will be applicable to wireless networks, too. The emphasis of this research proposal is on DoS attacks in general, and DDoS attacks are treated as a subset of DoS attacks. DDoS attacks are based on the same mechanisms as basic DoS attacks, but there is one exception during the deployment phase. A DDoS tool needs to be installed on many vulnerable hosts. The installation of DoS software on a single vulnerable host is, however, a common prerequisite for most DoS attacks. Thus attack and defense mechanisms described in this dissertation are applicable to both DoS and DDoS attacks.

VII. POSSIBLE OUTCOME

The main contributions of this proposed work include,

- 1) A comprehensive and well-structured description can be given about what DoS attacks really are? How DoS attacks can be carried out in IP networks? And how one can defend against DoS attacks in IP networks. A good understanding of existing attack mechanisms and available defense mechanisms is a prerequisite for succeeding in mitigating these attacks cost effectively.
- 2) An overview of an organized approach for selecting a comprehensive set of defense mechanisms against DoS attacks is given. This emphasizes the importance of basic security mechanisms at every host in the Internet, the importance of risk management in choosing additional defenses when basic defenses are not enough, and the necessity of implementing new defenses against such important DoS attacks for which there are no existing defenses.

3) A new defense mechanism for protecting organization-specific name servers will be described and simulated.

4) Since knowledge about DoS and DDoS is in primitive stage, we are hopeful to extend above objectives to study DoS attack in mobile ad hoc networks. An earlier attempt is found successful in some similar work [6].

CONCLUSION

This proposal aim to evaluate the DoS problems and availability of defence mechanism. It is understood that the existing defence mechanisms are mainly passive, in the sense that the target host or network is impaired before the attack source(s) can be found and controlled. We wish to propose a novel concept of active defence against DoS attacks by mitigating them in the Internet. This proposed style has sufficient advantages over conventional passive defence mechanisms. However, this is only the first step toward realizing the secure Internet paradigm. The proposed work can also be extended for designing of robust active defence architecture, developing a sensitive and accurate surveillance system, or for a powerful active trace back system and deployment of such system in real Internet environment.

REFERENCES

- [1] L. Zhou and Z. Haas. Securing ad hoc networks. IEEE Network, 13(6):24--30, November/December 1999.
- [2] Y. Zhang and W. Lee, "Intrusion detection in wireless ad hoc networks," ACM MOBICOM, 2000.
- [3] P.Papadimitratos and Z.J. Haas, "Secure Routing for Mobile Ad Hoc Networks," SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CND S 2002), San Antonio, TX, January 27-31, 2002.
- [4] S.Marti, T.Giuli, K.Lai and M.Baker, "Mitigating Routing Behavior in Mobile Ad Hoc Networks", *Proceedings of Mobicom 2001*, Rome,2001.
- [5] X Zeng, R. Bagrodia, and M. Gerla. *GloMoSim: a library for parallel simulation of large-scale wireless networks*. In Proceedings of the 12th Workshop on Parallel and Distributed Simulations, May 1998. 11.
- [6] Jean-Pierre Hubaux, Levente Buttyan, Srdjan Capkun, "The Quest for Security in Mobile Ad Hoc Networks", In Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC), Long Beach, CA, USA, October 2001.

Authors

Dr. S.D.Khamitkar: He is PhD in computer science and has 15+ research papers in International Conferences and journals. His interest area includes ICT, Green computing and Network Security.

P.U.Bhalchandra , N.K. Deshmukh , S.N.Lokhande : These are SET-NET qualified faculties and have 8+ years teaching experience. They have 5+ research papers in international conferences and journals. At present they are also working on research related to ICT and Green computing. The present paper is research topic of Mr. S.N.Lokhande

S.S.Phulari , R.P.Rathod : These are also faculties and have qualified M.Phil in computer science . They have 2+ papers in international conferences and journals.

An Efficient and Minimum Cost Topology Construction for Rural Wireless Mesh Networks

Prof. V. Anuratha & Dr. P. Sivaprakasam

Abstract

Many research efforts as well as deployments have chosen IEEE802.11 as a low-cost, long-distance access technology to bridge the digital divide. IEEE 802.11 Wi-Fi equipment based wireless mesh networks have recently been proposed as an inexpensive approach to connect far-flung rural areas. To establish such network high-gain directional antennas are used to achieve long-distance wireless point-to-point links. Some nodes in the network are called gateway nodes and are directly connected to the wired internet, and the remaining nodes connect to the gateway(s) using one or more hops.

In this paper the cost of constructing the antenna towers required is investigated. The problem is NP hard is shown and that a better than $O(\log n)$ approximation cannot be expected, where n is the number of vertices in the graph. To minimize the construction cost a new algorithm is proposed called constant time approximation algorithm.

The results of proposed approximation algorithm are compared with both the optimal solution, and a naive heuristic.

INTRODUCTION

There has been a huge proliferation of Internet and other communication based services in the last two decades. However, this spread is confined to developed countries, and metropolitan pockets of developing countries. This is really unfortunate for developing countries like India, where around 74% of the population

is rural and are on the wrong side of the digital divide.

Bridging this divide necessitates, providing internet connectivity to each and every village. Providing the same by expanding the current telephone network to rural areas is infeasible because of the huge initial infrastructure costs. Also, deployment of cellular wireless would not be sustainable because of its business model, which demands more high-paying consumer density.

Emerging technologies like 802.16 WMAN[12],[13], have not yet reached the scale of competitive mass production, hence the equipments are expensive. In this regard, the 802.11 Wi-Fi has shown tremendous growth and acceptance as a last hop access solution, because of their low price. Although 802.11 was primarily designed for indoor operation, but [3] has established the possibility of using 802.11 in long-distance networking.

The diverse requirements are in provisions of 1) Communication pattern which deals with the mode of communication one-to-one, one to-many, many-to-one, and many-to-many, 2) Delay (real-time, non-real-time, and delay-tolerant), 3) Service availability (centralized, distributed, and location-aware) that deals with the awareness of the availability of different services, such as Internet access, real-time communications, content distribution, interactive gaming, medical applications, and vehicular safety applications. 4) Security and 5) Reliability. An essential requirement to establish long-distance links is that line-of-sight is maintained between the radio antennas at the end-points.

To ensure line-of-sight across such long distances would require the antennas to be mounted on tall towers. The required height of the towers depends both on the length of the link, and the height of the obstructions along the link.

The cost of the tower depends on its height and the type of material used. For relatively short heights (10- 20 meters) antenna masts are sufficient. For greater heights, sturdier and much more expensive antenna towers are required. In this paper, several important contributions are made towards developing efficient algorithms [14] to solve this problem. First, the requirements to establish a point-to-point 802.11 link between two nodes of a given network graph is described. Then the formal definition of the Topology Construction problem (denote by TC) is given. It's proved that the problem to be NP hard by a reduction from the set-cover problem. The approximation algorithm is presented for this NP hard problem and the establishment cost of the tower in rural areas using constant time approximation algorithm is presented. The rest of this paper is organized as the following. Section II gives the Related Works of this technique is presented. In Section III, the Methodology of proposed approach is given. Section IV has the Experiment Results and this paper is concluded in Section V.

RELATED WORKS

802.11-based long-distance networks have been proposed as a cost-effective option to provide Internet connectivity to rural areas in developing regions, to enable Information and Communication Technology (ICT) services [5].

Rural areas (especially in developing regions) have populations with very low paying capacities. Hence, a major factor in network deployment is the cost of the infrastructure and the network equipment. In this context, efficient algorithms are investigated for the minimum cost topology construction problem in rural wireless mesh networks. R. Ramanathan *et.al*, [4] discussed

on the most critical design issues in multihop wireless networks. Topology control has been investigated extensively in the literature. Nevertheless, it is noted that most existing studies do not consider the requirements on upper layer applications or services. In this article the author address the topology control issues on service-oriented wireless mesh networks. In particular, the author provides a comprehensive survey of existing works on topology control from a service-oriented perspective. A general framework for topology control in service-oriented WMNs is proposed. To demonstrate the effectiveness of the framework, a case study is conducted in which the main objective is to maximize the overall throughput in a network with random unicast traffic. The performance of this topology control scheme is evaluated by numerical results. In addition, it is illustrated that the generated topology can support advanced technologies, including network coding and physical-layer network coding, which can significantly improve the throughput capacity of a network. The cost of laying wire to rural areas is prohibitively expensive.

Also, traditional wireless technologies such as cellular data networks (e.g., EV-DO) and upcoming technologies like IEEE 802.16 WiMAX have prohibitively expensive equipment costs. As a result, there has been considerable recent interest [6], [7], [8] in the design of rural mesh networks using IEEE 802.11 (Wi-Fi) equipment. The cost of an 802.11 radio (»\$50/PCMCIA card) is orders of magnitude less than that of cellular/WiMAX base stations. Thus, this approach is an attractive option for building low cost networks. D. S. Lun *et.al*, [9] presented a distributed random linear network coding approach for transmission and compression of information in general multisource multicast networks.

Network nodes independently and randomly select linear mappings from inputs onto output links over some field. The author shows that this achieves capacity with

probability exponentially approaching 1 with the code length. Random linear coding is demonstrated which performs compression when necessary in a network, generalizing error exponents for linear Slepian-Wolf coding in a natural way. Benefits of this approach are decentralized operation and robustness to network changes or link failures. The author shows that this approach can take advantage of redundant network capacity for improved success probability and robustness. Some potential advantages of random linear network coding are illustrated over routing in two examples of practical scenarios: distributed network operation and networks with dynamically varying connections. The derivation result also yields a new bound on required field size for centralized network coding on general multicast networks.

METHODOLOGY

In this paper a Novel Topology Control Scheme is used to identify a set of semi-permanent highways, such that the best throughput capacity of the network can be obtained. Particularly, the wireless highways are predicted to be rather similar to the highway system in public transportation system, which can efficiently provide connectivity in real application.

A. Computing tower heights at the end-points of a link

Consider two nodes, u and v that are separated by a distance l_{uv} . The edge (u, v) is considered to be covered if an 802.11 based point-to-point communication link can be established between u and v . Assume that the transmit powers [15] and the gains of the antennas at both ends are sufficient to overcome the free-space path loss between the two points. The first basic requirement to cover the edge between u and v is that there is a clear visual line-of-sight between the antennas at the end-points (as shown in Fig. 4a). In other words, the line joining the antennas mounted on the towers should clear any obstructions along the path. Secondly, it is also required that RF line-of-sight is maintained between the two points. This is

determined by an elliptical area between u and v termed the first Fresnel zone. To establish RF line-of-sight, a significant area of the Fresnel zone ($> 60\%$ of the radius of the Fresnel zone at the location of the obstruction [1]) should also clear all obstructions between u and v . However, this can be simply modeled by extending the height of the obstruction to include the radius of the Fresnel zone that has to be in the clear.

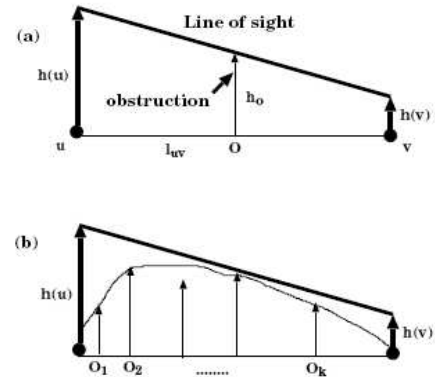


Figure1: Computing the height of towers at the end-points of a link

In reality, there can be multiple obstructions between u and v . As in Figure 4b, consider multiple obstructions, O_1, O_2, \dots, O_k between u and v . Now, let $h(u)$ and $h(v)$ represent the tower heights at the nodes of u and v . Covering edge (u, v) requires a visual and RF line-of-sight connection between the towers at its two terminal nodes. This would imply that the straight line fu_v joining the top of the two towers (of heights $h(u)$ at u and $h(v)$ at v) should clear every obstruction in (u, v) . Hence, it is noted that given a particular pair of tower heights at u and v , deciding whether these heights covers edge (u, v) can be done in time linear in the number of obstructions on that edge.2

B. Modeling tower costs

An important component in this problem is the nature of the cost function that maps tower heights to the cost of building the tower.

There are two types of antenna towers that are used. For heights less than 20 meters, one can use the cheaper masts. For greater heights, the more expensive steel towers is used.

Further, there is an order of magnitude difference between the cost of the cheaper masts and that of the steel towers. Thus, roughly speaking, the cost function is constant as long as the cheaper masts can be used and becomes linear in height once the steel towers are needed, with a jump in cost when we switch from masts to steel towers.

Let us denote the height at which the material of the tower has to be switched as h_{min} .

Further, there is a physical restriction on the maximum possible height of a tower, denoted by h_{max} . Thus, the cost function c can be formally defined as

$$c(h) = \begin{cases} K & \text{if } 0 \leq h \leq h_{min} \\ A_h + B & \text{if } h_{min} < h \leq h_{max} \end{cases}$$

Where A , B and K are constants and $Ah_{min} + B \gg K$. Although, in practice, the cost function can be modeled as discussed above, our algorithm works with a much more general cost functions. Specifically, we only require the cost function c to satisfy the following two natural properties C1 and C2. C1

Given the tower costs at two neighboring nodes u and v , it can be determined (in polynomial time) whether the corresponding tower heights cover the edge (u,v) . This simply requires that the corresponding tower height can be computed (in polynomial time) given the tower cost. As mentioned earlier, determining whether the height of the towers is sufficient to cover an edge can be done in polynomial time. C2 the cost function is monotonically increasing with height, i.e., $h_1 \geq h_2$ and $c(h_1) \geq c(h_2)$ for any values of h_1 and h_2 .

It is easy to verify that the cost function c defined earlier in this section satisfies both

of the above properties. In the remainder of this paper, when the height function h is unambiguous, often the cost of the tower is denoted at a node v as $c(v)$ rather than $c(h(v))$.

C. The Topology Control Scheme

In this framework, it is considered that the highways can be partitioned into two groups, horizontal and vertical. Highways in each group can operate simultaneously because they are mutually parallel and can be placed away enough to reduce interference below a certain threshold. Consequently, horizontal and vertical highways will partition the whole geographical area into grids, in which nodes will try to forward their traffic to the nodes on neighboring highways. The combination of the following parameters can be considered for the Topology Control Scheme

- 1) **Transmission range:** Transmission range of each node in the network is traditionally an important design parameter in topology control. In general, a smaller transmission range will improve the channel reuse but may compromise the connectivity. A larger transmission range will improve the connectivity but reduce the channel reuse. Therefore, an appropriate range is chosen as a trade-off between connectivity and channel reuse.
- 2) **Type of antenna:** When directional antenna or beam forming is used it may improve the capacity of the network by reducing the interference and improve the transmission quality.
- 3) **Traffic pattern:** Traffic pattern is very important parameter to the topology. In most studies previously done suggest that the traffic is broadcast. With such an assumption, the problem is formulated in a way such that the overall transmission for each

message is minimized. However, broadcast traffic may only be a special case in the future service-oriented WMN, in which a variety of patterns may appear, from one-to-one to many-to-many [10].

- 4) **Quality of service (QoS):** To achieve an efficient network a crucial issue is to enable services with certain QoS requirements, such as bandwidth, delay, security and reliability.

For the first step of this study, only omnidirectional antenna and purely random unicast traffic pattern are considered. Moreover, the random wireless network elaborated to gain insights for the future investigation.

Figure 2: Approximation algorithm for NP-hard problem

Constant Time Approximation Algorithm

1. $F \leftarrow \emptyset$ *Comment: Implicitly set growth variables $y_s \leftarrow 0$ and cumulative growth variables $w(s) \leftarrow 0$ for all $s \in V$. $w(s)$ denotes the total potential expended in component S including any sub components that were merged in creating S . Also implicitly set $\pi_v \leftarrow \frac{L}{(1-\alpha)k}$ for all node $v \neq r$.*
2. $c \leftarrow \{v : v \in V, v \neq r\}$.
For each $v \neq r$ set $d(v) \leftarrow$
3. *0. Comment: $d(v)$ denotes the distance of node v to the boundary of the component containing v .*
For each $v \in V$ if $v = r$ then $\lambda(\{v\}) \leftarrow 0$ else $\lambda(\{v\}) \leftarrow$
4. *1. Comment: $\lambda(s)$ equal 1 if s active and 0 otherwise.*
5. *while there are active components*
6. *Comment: Find the next event*
7. *Find edge $e = (i, j)$ with $i \in C_p \in C, j \in C_q \in C, C_p \neq C_q$ that minimizes*
$$\epsilon = \frac{c_e - d(i) - d(j)}{\lambda(C_p) + \lambda(C_q)}$$
8. *Find $\tilde{C} \in C$ with $\lambda(\tilde{C}) = 1$ that minimizes $\epsilon_2 = |\tilde{C}| \cdot \frac{L}{(1-\alpha)k} - \omega(\tilde{C})$*
9. $\epsilon = \min(\epsilon_1, \epsilon_2)$
10. $\omega(C) \leftarrow \omega(C) + \epsilon \cdot \lambda(C)$ *for all $C \in c$*
11. *for all $v \in C_p \in c$*
12. $d(v) \leftarrow d(v) + \epsilon \cdot \lambda(C_p)$
13. *if $\epsilon = \epsilon_2$ (i.e., \tilde{C} is deactivated before C_p and C_q meet)*
14. $\lambda(\tilde{C}) \leftarrow 0$
15. *Mark all unlabeled vertices of \tilde{C} with label \tilde{C} . else (i.e., C_p and C_q meet before \tilde{C} is deactivated)*
16. $F \leftarrow F \cup \{e\}$
17. $C \leftarrow C \cup \{C_p \cup C_q\} - \{C_p\} - \{C_q\}$.
18. $\omega(C_p \cup C_q) \leftarrow \omega(C_p) + \omega(C_q)$
19. *if $r \in C_p \cup C_q$ then $\lambda(C_p \cup C_q) \leftarrow 0$ else $\lambda(C_p \cup C_q) \leftarrow 1$.*
20. *for every unlabeled vertex $v \in C$,*

EXPERIMENTAL RESULTS

In this section, extensive numerical simulations are carried out to evaluate our approximation algorithm with the optimal solution and also a naive heuristic. For this simulations, synthetic topologies are generated that aim to match the geographical structure of village clusters. Simulation setup is described in more detail as follows.

A. Generating synthetic graph topologies.

A circular plane is considered with a radius of 25Kms. Nodes are placed at random locations on this plane. A link (u,v) is considered between any two nodes, u and v , and for these simulations, assume just one obstacle, ouv , located on the middle of this link.

The height of the obstruction (ho) is selected randomly with a maximum value of 20 meters - the typical height of trees and small houses in a rural setting. A weight wuv is assigned to the link equal to twice the effective height of the obstruction on this link. As described earlier, the effective height of an obstruction, is the sum of the physical height (ho) and 60% of rf , the radius of the fresnel zone.

B. Naive heuristic

In order to compare with the proposed approximation algorithms, a naive heuristic is described for selecting connected subgraphs and assigning heights to the nodes. As a first step, to select a connected subgraph of an input graph G , the heuristic computes the minimum spanning tree (MST), T of G (using the link weights computed as described above). Next, the heuristic has to assign heights to the nodes in G , so as to cover all the edges in T while minimizing the total cost.

Given a set of links to be covered, the height assignment problem is formulated as a simple LP, and compute the heights required on every node.

C. Comparing with the naive heuristic

Now the naive heuristic described above is compared with the proposed approximation algorithm. Graphs considered with number of nodes $n = 10, 15, 20, \dots, 50$. For each value of n we generate 50 graph instances. For each graph, C_{naive} the cost of the solution produced by the naive heuristic, and C_{approx} cost of the approximation algorithm is computed. With this value it is observed that the proposed approximation algorithm performs substantially better than the naive heuristic.

On average, the solutions returned by the naive heuristic range from 60% (for $n = 10$) to as much as 225% more expensive (for $n = 50$) compared to the solution returned by the naive algorithm.

D. Comparing with the optimal solution

The optimal solution is computed by solving an ILP that models the topology construction problem. The CPLEX LP-solver [11] to solve this ILP. This approach is, however, computationally very expensive, and the LP-solver could return solutions for graphs with at most 11 nodes.

The solution returned by our approximation algorithm is compared with the optimal solution for graphs with number of nodes $n = 8, 9, 10, 11$. For each value of n we generate 50 graphs. For each graph C_{approx} , the cost of the solution returned by proposed approximation algorithm, and C_{opt} , the cost of the optimal solution.

Then the mean and standard deviation is computed over all graphs for different values of n $R_{opt} = \frac{C_{approx} - C_{opt}}{C_{opt}}$.

n	Mean (std. dev.) of Ropt
8	0.58 (0.30)
9	0.57 (0.25)
10	0.55 (0.23)
11	0.52 (0.25)

Table1: New Approximation algorithm vs. Optimal Solution.

The results presented in Table 7 show that the incorporated approximation algorithm gives solutions that are 50 - 60% more expensive than the optimal solution (for small values of n). Thus, the new approximation algorithm performs much better than the worst case guarantee of $O(\log n)$ on the approximation factor.

While this gap between constant time approximation algorithm and the optimal solution is not small, its expected computing the optimal solution (even if it has to be done only once) is practically infeasible for real-life networks. Moreover, this approximation algorithm performs substantially better in practice than the naive heuristic and previous approximation algorithms

CONCLUSION

In this paper an overview of establishing a low cost wireless mesh network for rural areas is presented. In rural areas nodes are connected using long distance 802.11 wireless links which are established using high-gain directional antennae. The main problem is with the topology construction for long distance wireless communication.

An efficient approximation algorithm is proposed for the topology construction problem in rural mesh networks. This work introduces a number of

open research problems in the topology construction.

One immediate problem is to consider the case of $k \geq 2$ vertex or edge connectivity, similar to the power optimal network construction for k -connectivity. Another important research direction is the geometric version of this problem. In practice, all nodes within a certain distance of each other can establish a link.

In this paper, the location of the towers is assumed to be fixed (within a village). A variant of the problem would make the location of the tower to be a variable.

This method has added flexibility than the previous method which would result in reduced cost. The numerical experiments demonstrate that the proposed method with constant time approximation algorithm performs well within its worst case performance bounds, and outperforms the naive heuristic by a substantial margin.

References

- [1] I. Akyildiz and X. Wang, "A survey on wireless mesh networks," IEEE Communications Magazine, vol. 43, no. 9, pp. S23–S30, Sept. 2005.
- [2] M. Lee, J. Zheng, Y.-B. Ko, and D. Shrestha, "Emerging standards for wireless mesh technology," IEEE Wireless Communications, vol. 13, no. 2, pp. 56–63, April 2006.
- [3] Pravin Bhagwat, Bhaskaran Raman, and Dheeraj Sanghi, "Turning 802.11 Inside-Out", In HotNets-II, Nov 2003.
- [4] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit poweradjustment," vol. 2, 2000.
- [5] Eric Brewer, Michael Demmer, Bowei Du, Kevin Fall, Melissa Ho, Matthew Kam, Sergiu Nedeveschi, Joyojeet Pal, Rabin Patra, and Sonesh Surana. The Case for Technology for Developing

- Regions. IEEE Computer, 38(6):25–38, June 2005.
- [6] P. Dutta, S. Jaiswal, K. Naidu, D. Panigrahi, R. Rastogi, and A. Todimala. Villagenet: A low-cost, 802.11-based mesh network for rural regions. In Wireless Systems: Advanced Research and Development Workshop (WISARD), 2007.
- [7] R. Patra, S. Nedeveschi, S. Surana, A. Sheth, L. Subramanian, and E. Brewer. WiLDNet: Design and Implementation of high performance wifi based long distance networks. In NSDI, 2007.
- [8] B. Raman and K. Chebrolu. Design and evaluation of a new MAC for long distance 802.11 mesh networks. In Mobicom, 2005.
- [9] D. S. Lun, N. Ratnakar, M. Medard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, “Minimum-cost multicast over coded packet networks,” IEEE Transaction on Information Theory, vol. 52, no. 6, pp. 2608–2623, June 2006.
- [10] K. Lu, Y. Qian, H.-H. Chen, and S. Fu, “WiMAX Networks: From Access To Service Platform,” IEEE Network, 2008, accepted.
- [11] ilogcplex.
<http://www.ilog.com/products/cplex/>.
- [12] M. T. Hajiaghayi, N. Immorlica, and V. S. Mirrokni. Power optimization in fault-tolerant topology control algorithms for wireless multi-hop networks. In MOBICOM, 2003.
- [13] M. T. Hajiaghayi, G. Kortsarz, V. S. Mirrokni, and Z. Nutov. Power optimization for connectivity problems. In IPCO, 2005.
- [14] Z. Nutov. Approximating minimum power covers of intersecting families and directed connectivity problems. In APPROX-RANDOM, 2006.
- [15] S. Sen and B. Raman, “Long distance wireless mesh network planning: Problem formulation and solution”, In WWW, 2007.
- [16] P. N. Klein and R. Ravi. A nearly best-possible approximation algorithm for node-weighted steiner trees. J. Algorithms, 19(1), 1995.

AUTHOR BIOGRAPHIES

Mrs. V. Anuradha graduated with B.Sc Computer Science in the year 1995 and completed M.C.A at Madras University in the Year 2000. Completed her M.Phil in the year 2003 and also got Guide Approval for M.Phil in Bharathiar University, Peiyar University, Bharathiar University and currently doing her Ph.d.Hear area of interest is Networks and data mining. Mrs. V. Anuratha have guided 20 M.Phil scholars and she have participated and presented many papers in the national and international conferences and etc..

Currently she is working as a H.O.D – PG Department of Computer Science at Sree Saraswathi Thyagaraja College with a decade a teaching experience

Co author biography.....

Dr. P. Sivaprakasam :

He have completed his M.Sc(c.s) in the year 1986, M.Phil in the year 1995 and Ph.d in 2005 on the topic “An Analysis of Web performance and caching”. He have 19 years of teaching experience. He have published 6 papers at national level and 3 at international level. He was also sanctioned with 2 UGC research projects.

He is now currently working as a Associate professor in Computer Science at Sri Vasavi college of Arts and Science. His areas of interest are Internet, Computer Networks, Service Oriented Architecture.

Reinforcement Learning by Comparing Immediate Reward

Punit Pandey

Department Of Computer Science and Engineering,
Jaypee University Of Engineering And Technology

Deepshikha Pandey

Department Of Computer Science and Engineering,
Jaypee University Of Engineering And Technology

Dr. Shishir Kumar

Department Of Computer Science and Engineering,
Jaypee University Of Engineering And Technology

Abstract— This paper introduces an approach to Reinforcement Learning Algorithm by comparing their immediate rewards using a variation of Q-Learning algorithm. Unlike the conventional Q-Learning, the proposed algorithm compares current reward with immediate reward of past move and work accordingly. Relative reward based Q-learning is an approach towards interactive learning. Q-Learning is a model free reinforcement learning method that used to learn the agents. It is observed that under normal circumstances algorithm take more episodes to reach optimal Q-value due to its normal reward or sometime negative reward. In this new form of algorithm agents select only those actions which have a higher immediate reward signal in comparison to previous one. The contribution of this article is the presentation of new Q-Learning Algorithm in order to maximize the performance of algorithm and reduce the number of episode required to reach optimal Q-value. Effectiveness of proposed algorithm is simulated in a 20 x20 Grid world deterministic environment and the result for the two forms of Q-Learning Algorithms is given.

Keywords-component; Reinforcement Learning, Q-Learning Method, Relative Reward, Relative Q-Learning Method.

I. INTRODUCTION

Q-Learning algorithm proposed by Watkins [2,4] is a model free and online reinforcement learning algorithm. In reinforcement learning selection of an action is based on the value of its state using some form of updating rule. There is an interaction between agent and environment where the agent has to go through numerous trials in order to find out the best action. An agent chooses that action which has maximum reward obtained from its environment. The reward signal may be positive or negative depends on the environment.

Q-learning has been used in many applications because it does not require the model of environment and is easy to implement. State-action value, a value for each action from each state, converges to the optimal value as state-action pairs are visited many times by the agent.

In this article we propose a new relative reward strategy for agent learning. Two different form of Q-Learning method is considered here as a part of study. First form of Q-Learning method uses a normal reward signal. In this algorithm Q-value evaluates whether things have gotten better or worse than

expected as a result of an action selection in the previous state. The action selected by agents is most favorable which has lower TD error. Temporal difference is computed on the basis of normal reward gain by agents from its surroundings. An estimated Q-value in the current state is then determined using Temporal Difference. Agent actions are generated using the maximum Q-values. The second form of Q-Learning algorithm is an extension towards a relative reward. This form of Q-Learning method utilizes the relative reward approach to improve the learning capability of algorithm and decreases the number of iteration. In this algorithm only those action is selected which has a better reward from its previous one.

This idea comes from psychological point of views that human beings tend to select only those action which has higher reward value. However, this algorithm is not suitable for multi agent problems. To demonstrate effectiveness of the proposed Q-Learning algorithm, java applet is utilized to simulate a robot that reaches to a fixed goal. Simulation result confirms that the performance of proposed algorithm is convincingly better than conventional Q-learning.

This paper is organized as follows: Basic concept of reinforcement learning is presented in section 2. Section 3 describes about the conventional Q-Learning method. Section 4 presents a new Relative Q-Learning in context of relative immediate reward. Section 5 describes Experimental setup & results and concluding remarks follow in Section 6.

II. REINFORCEMENT LEARNING

Reinforcement learning (RL) is a goal directed learning methodology that is used to learn the agents. In Reinforcement learning [1,5,6,7,8,9] the algorithm decide what to do and how to map situations to actions so that we maximize a numerical reward signal. The learner is not advised which actions to take, but instead it discover which actions provide the maximum reward signal by trying them. Reinforcement learning is defined by characterizing a learning problem. Any algorithm that can able to solve the defined problem, we consider to be a reinforcement learning algorithm. The key feature of reinforcement learning is that it explicitly considers the whole problem of a goal-directed agent interacting with an uncertain environment. All reinforcement learning agents [3,10,11,12] have explicit goals, can sense aspects of their environments, and can choose actions to influence their environments. In

reinforcement learning agent prefer to choose actions that it has tried in the past and found to be effective in producing maximum reward. The agent has to exploit based on what it already knows in order to obtain reward and at the same time it also has to explore in order to make better action selections in the future. Reinforcement learning has four elements policy, reward function, value function and model of environment.

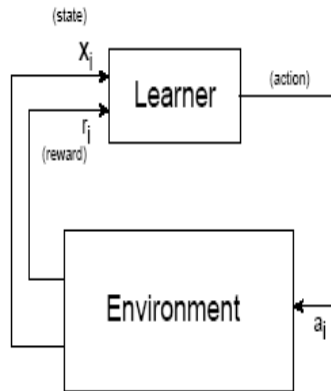
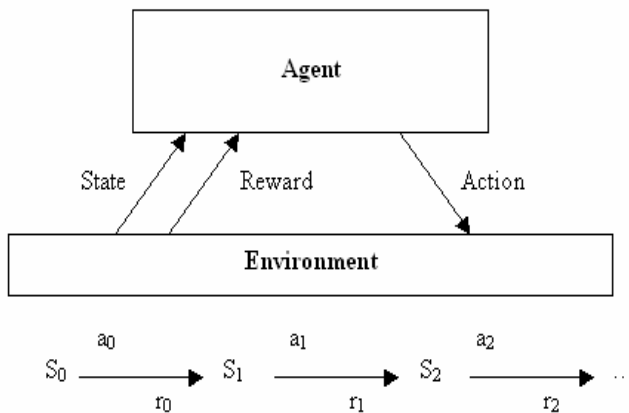


Figure 1.1: Reinforcement learning



Goal: learn to choose actions that maximize:
 $r_0 + \gamma r_1 + \gamma^2 r_2 + \dots$, where $0 \leq \gamma < 1$

Figure 1.2: Reinforcement learning

Model of the environment is an optional element because reinforcement learning also supports the model free algorithms like Q-learning.

A policy for our agent is a specification of what action to take for every Input. In some cases policy may be a simple function or look-up table or sometime it can be an extensive computation. The policy is the core of reinforcement learning

agent because it alone is sufficient to take the decision on further action.

III. Q-LEARNING

Q-learning is a form of model-free reinforcement learning [2] (i.e. agent does not need an internal model of environment to work with it). Since Q-learning is an active reinforcement technique, it generates and improves the agent's policy on the fly. The Q-learning algorithm works by estimating the values of state-action pairs.

The purpose of Q-learning is to generate the Q-table, $Q(s,a)$, which uses state-action pairs to index a Q-value, or expected utility of that pair. The Q-value is defined as the expected discounted future reward of taking action a in state s , assuming the agent continues to follow the optimal policy. For every possible state, every possible action is assigned a value which is a function of both the immediate reward for taking that action and the expected reward in the future based on the new state that is the result of taking that action. This is expressed by the one-step Q-update equation [2,4,10,13,14].

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (1)$$

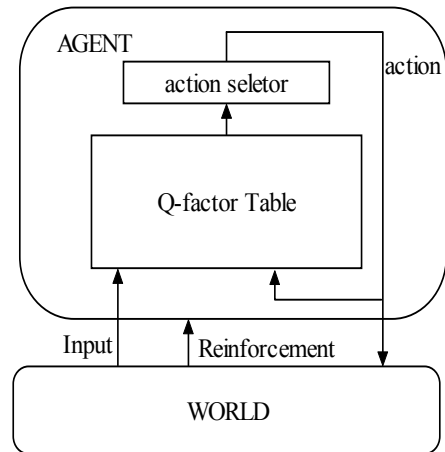


Figure 2: Structure of the Q-Learning agent

Where α is the learning factor and γ is the discount factor. These values are positive decimals less than 1 and are set through experimentation to affect the rate at which the agent attempts to learn the environment. The variables s and a represent the current state and action of the agent, r is the reward from performing s' and a' , the previous state and action, respectively.

The discount factor makes rewards earned earlier more valuable than those received later. This method learns the values of all actions, rather than just finding the optimal policy. This knowledge is expensive in terms of the amount of information that has to be stored, but it does bring benefits. Q-learning is exploration insensitive, any action can be carried out at any time and information is gained from this experience. The agent receives reinforcement or reward from the world,

and returns an action to the world round and round as shown below:

A. Elementary parts of Q-learning:

Environment:

Q-learning based on model-free mode of behavior i.e the environment is continuously changing. Agent does need to predict future state. Environment can be either deterministic or non-deterministic. In deterministic environment application of single state lead to a single state where as in nondeterministic environment application of a single action may lead to a number of possible successor states. In case of non-deterministic environment, each action not only labeled with expected immediate reward but also with the probability of performing that action. For the sake of simplicity we are considering deterministic environment in this thesis work.

Reward Function:

A reward function defines the goal in a reinforcement learning problem. it maps each perceived state (or state-action pair) of the environment to a single number, a reward, indicating the intrinsic desirability of that state. A reinforcement learning agent's sole objective is to maximize the total reward it receives in the long run. The reward function defines what the good and bad events are for the agent.

Action-value function:

The Q-learning learning is based upon Quality-values (Q-values) $Q(s,a)$ for each pair (s,a) . The agent must cease interacting with the world while it runs through this loop until a satisfactory policy is found. Fortunately, we can still learn from this. In Q-learning we cannot update directly from the transition probabilities-we can only update from individual experiences. In 1 step Q-learning, after each experience, we observe state s' , receive reward r , and update:

$$Q(s, a) = r + \gamma \max_{a'} Q(s', a') \quad (2)$$

B. Q-learning Algorithm:

Initialize $Q(s, a)$ arbitrarily

Repeat (for each episode)

Choose a starting state, s

Repeat (for each step of episode):

Choose a from s using policy derived from Q

Take action a , observe a immediate reward r , next state s'

$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$
 $s \leftarrow s'$;

Until state s' match with the Goal State

Until a desired number of episodes terminated

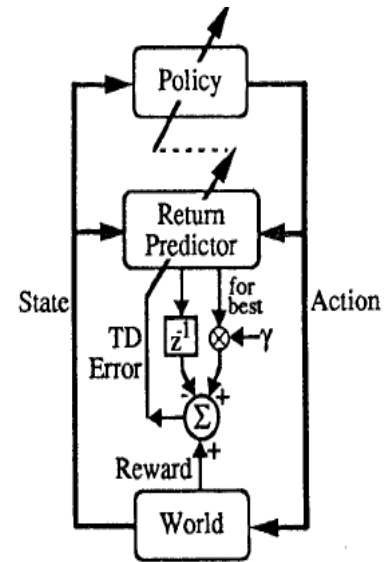


Figure 3: Q-Learning Architecture

IV. RELATIVE Q-LEARNING

This section introduces a new approach Relative reward to conventional Q-learning that makes Relative Q-Learning. Conventional Q-learning has been shown to converge to the optimal policy if the environment is sampled infinitely by performing a set of actions in the states of the environment under a set of constraints on the learning rate α . No bounds have been proven on the time of convergence of the Q-learning algorithm and the selection of the next action is done randomly when performing the update. This simply mean that the algorithm would take a longer time to converge as a random set of states are observed which may or may not bring the state closer to the goal state. Furthermore, it means that this function cannot be used for actually performing the actions until it has converged as it has a high chance of not having the right value as it may not have explored the correct states. This is especially a problem for environments with larger state spaces. It is difficult to explore the entire space in a random fashion in a computationally feasible manner. So by applying below mention method and algorithm we try to keep the Q-learning algorithm near to its goal in less time and less number of Episode.

A. Relative Reward

Relative reward is a concept that compares (current reward with the previous received reward) two immediate rewards. The objective of the learner is to choose actions maximizing discounted cumulative rewards over time. Let there is an agent in state s_t at time t , and assume that he chooses action a_t . The immediate result is a reward r_t received by the agent and the state changes to s_{t+1} . The total discounted reward [2,4] received by the agent starting at time t is given by:

$$r(t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots + \gamma^n r_{t+n} + \dots \quad (3)$$

Where γ is discount factor in the range of (0:1).

The immediate reward is based upon the action or move taken by an agent to reach the defined goal in each episode. The total discounted reward can maximize in less number of episode if we select the higher immediate reward signal from previous.

B. Relative Reward based Q-Learning Algorithm

Relative reward based Q-learning is an approach towards maximizing the total discounted rewards. In this form of Q-learning we selected the maximum immediate reward signal by comparing it with previous one. This is expressed by the new Q-update equation.

$$Q(s, a) = Q(s, a) + \alpha [\max(r(s,a), r(s',a')) + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Algorithm:

Initialize $Q(s, a)$ arbitrarily

Repeat (for each episode)

Choose a starting state, s

Repeat (for each step of episode):

Choose a from s using policy derived from Q

Take action a , observe a immediate reward r , and next state s'

$$Q(s, a) = Q(s, a) + \alpha [\max(r(s,a), r(s',a')) + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

$s \leftarrow s'$;

Until state s' match with the Goal State

Until a desired number of episodes terminated

V. EXPERIMENTS & RESULTS

The Proposed Relative Q-Learning was tested on 10 x 10 and 20 x 20 grid world environment. In the Grid World Square There are four possible actions for the agent as it is a deterministic environment given in figure 4.

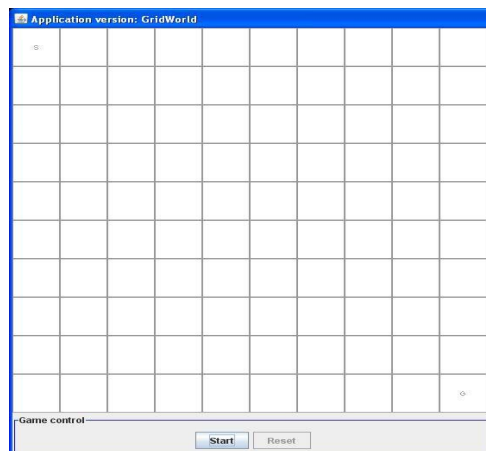


Figure4: A 10 x 10 Grid World Environment

In order to consider the situation of encountering a wall, the agent has no possibility of moving all the way in the given direction. When the agent enters into goal states, it receives 50 as a reward. We are also providing the immediate reward value by incrementing or decrementing the Q-value marked with S represent the start state and G represent the goal state. The purpose of the agent is to find out the optimum path to arrive at the goal state starting from the start state, and to maximize the reward it receives.

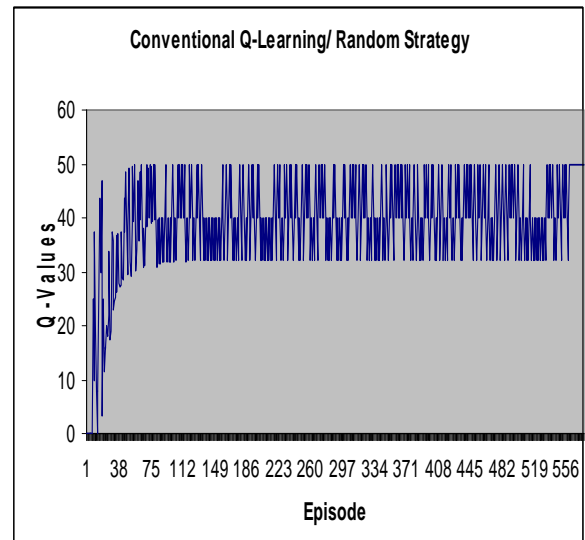


Figure5: Conventional Q-Learning.

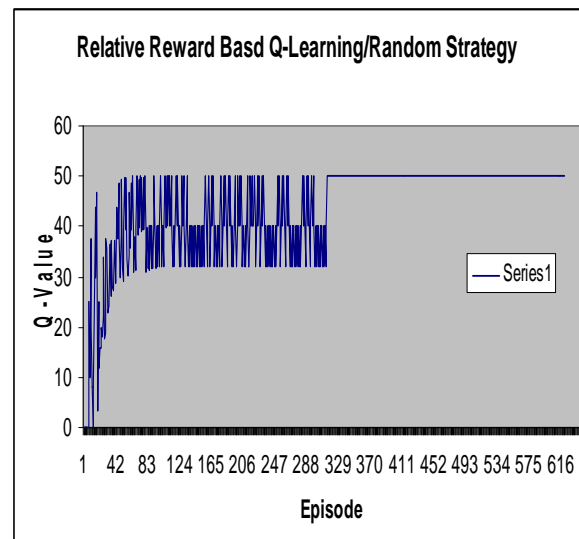


Figure6: Relative Q-Learning

We have executed 500 episodes to converge the Q-value. The grid world is a deterministic environment so the value of learning α and discount rate γ were set to 0.8. Figure 5 &

Figure 6 shows the relationship between Q-Values and the number of episode where x axis represents the number of episode and y axis represents the Q-values. Figure 5 represents the result of conventional Q-Learning where we can see that Q-value converges after executing 500 episodes but in figure 6 Relative Q-learning takes 300 episode

So we can say that convergence rate of relative Q-learning is faster than conventional Q-learning.

VI. CONCLUSION & FUTURE WORK

This paper proposed an algorithm which compares the immediate reward signal with its previous one. The agent will immediately return back to previous state if it will receive the lower reward signal for that particular move. If conventional Q-learning was applied in the real experiment, a lot of iterations were required to reach the optimal Q values. The Relative Q-learning algorithm was proposed for environment which used small amount of episodes to reach the convergence of Q-values. This new concept allows the agent to learn uniformly and helps in such a way so that it will not deviate from its goal. Part of future work may be included to verify the proposed algorithm in nondeterministic environment.

REFERENCES

- [1] J.F. Peters, C. Henry, S. Ramanna, Reinforcement learning with pattern-based rewards. in proceeding of forth International IASTED Conference. Computational Intelligence (CI 2005) Calgary, Alberta, Canada, 4-6 July 2005, 267-272
- [2] Technical Note Q-Learning Christopher J.C.H. Watkins and Peter Dayan Centre for Cognitive Science, University of Edinburgh, Scotland Machine Learning, 8, 279-292 (1992)
- [3] J.F. Peters, C. Henry, S. Ramanna, Rough Ethograms: Study of Intelligent System Behavior. In: M.A. Klopotek, S. Wierzbicki, K. Trojanowski (Eds), New Trends in Intelligent Information Processing and Web Mining (IIS05), Gdańsk, Poland, June 13-16 (2005), 117-126.
- [4] C. Watkins, "Learning from Delayed Rewards", PhD thesis, Cambridge University, Cambridge, England, 1989
- [5] J.F. Peters, K.S. Patnaik, P.K. Pandey, D. Tiwari, "Effect of temperature on swarms that learn", In Proceeding of IASCIT-2007, Hyderabad, INDIA
- [6] P.K. Pandey, D. Tiwari, "Temperature variation on Q-Learning", In Proceeding of RAIT in FEB 2008, ISM Dhanbad
- [7] P.K. Pandey, D. Tiwari, "Temperature variation on Rough Actor-Critic Algorithm", Global Journal Computer Science and Technology, Vol 9, No 4 (2009), Pennsylvania Digital Library
- [8] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: A survey Journal of Artificial Intelligence Research, 4, 1996, 237-285.
- [9] R.S. Sutton, A.G. Barto, and Reinforcement Learning: An Introduction (Cambridge, MA: The MIT Press, 1998).
- [10] C. Gaskett, Q-Learning for Robot Control. Ph.D. Thesis, Supervisor: A. Zelinsky, Department of Systems Engineering, The Australian National University, 2002.
- [11] Thrun. S. and Schwartz. A. (1993), Issues in using function approximation for reinforcement learning, in Proceeding of the 1993 Connectionist Models Summer School, Erlbaum Associates. Nj.
- [12] Richard S. Sutton, Reinforcement Learning Architectures, GTE Laboratories Incorporated, Waltham, MA 02254.
- [13] Tom O'Neill, Leland Aldridge, Harry Glaser, Q-Learning and Collection Agents, Dept. of Computer Science, University of Rochester
- [14] Vanden Berghen Frank, Q-Learning, IRIDIA, Universit Libre de Bruxelles

Information realization with statistical predictive inferences and coding form

D.Mukherjee

Sir Padampat Singhania University
Udaipur-313601,Rajasthan,India

P.Chakrabarti* , A.Khanna , V.Gupta

Sir Padampat Singhania University
Udaipur-313601,Rajasthan,India

Abstract—The paper deals with information realization in case of grid topology. Nodal communication strategies with clusters has also been cited. Information prediction has been pointed out with relevant statistical method, forward sensing, backward sensing and cumulative frequency form. Binary tree classifier theory has been applied for information grouping. The paper also deals with comparison analysis of information coding.

Keywords- grid topology ,forward sensing , backward sensing, binary tree classifier, information coding

I. INFORMATION MERGING IN GRID IN DIAGONAL APPROACH

In order to solve complex problems in artificial intelligence one needs both large amount of knowledge & some mechanisms for manipulating that knowledge to create solutions to new problems .Basically knowledge is a mapping of different facts with help appropriate functions for e.g. Earth is a planet. Can be realized as a function – *planet (Earth)*.

Information merging can be realized as combining different pieces of information to arrive at a conclusion. The different information elements can be related in different ways i.e. either in hierarchy or in form of a graph or even a mesh. Consider a mesh of size $m \times n$ i.e. m rows & n columns then if each intersection point has a information element placed on it then one way of merging element A with B can be covering a path of length $(5 \times N)$ (here $m=8$ & $n=9$). If weight of covering each path is considered same then in case of diagonal approach we can find a path of diagonal nature of length $5\sqrt{2}$ and then travelling a length $(N-5)$ in linear fashion thus finding a shortest path the same can also be determined by graph algorithms like Dijkstra's or kruskal's algorithm for minimum spanning tree. If each path is considered to be of zero weight then interestingly there is no sense travelling a path from A to B i.e. we can directly merge the two points i.e. we take point A & directly merge it with point B in such a case we need to have some stack like mechanism to determine the order in which the nodes arrive & are merged.

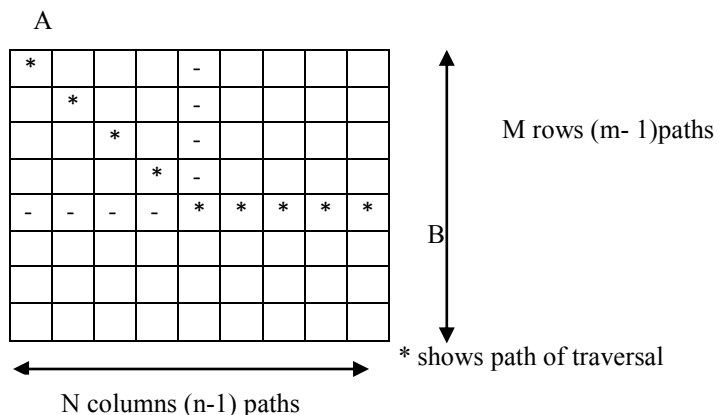


Fig1: Information merging in mesh/grid

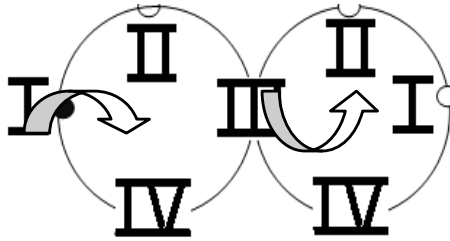
The above concept can be realized in DDM(Distributed Data Mining) where large amount of geographically scattered knowledge is merged & is mined to derive conclusions & make decisions for e.g. GIS i.e. the Geographical Information System which uses cartography(art of making maps) with various information elements(sources) to derive decision support results like which route to choose for a given destination.

II. INFORMATION MERGING IN CLUSTER NETWORKS

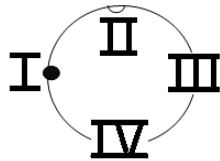
This section mainly focuses on the nodal communication between the farthest node in a $N \times N$ structure[1] and information realization indicates nodal message . Let us assume each cluster to be consisting of 16 nodes and then try to communicate between the source and the destination node as described in the fig1. The point to be noted here is that to establish the communication link between the adjacent elements or units of the cluster we have to have the communication in just reverse order in the 2 adjacent elements. The order of the communication is



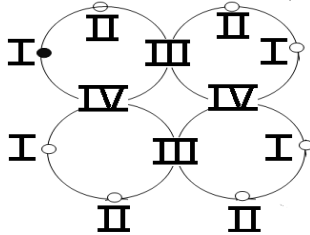
The condition can be visually imagined as follows:



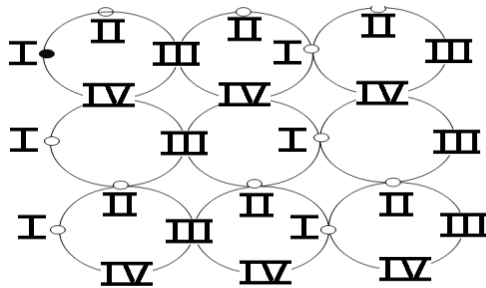
Now let us first talk about the case when there is only one element i.e., 1×1 . In this particular case if we want to communicate between the farthest node then there will be only 1 node in between the source and the destination which can be further visualized as follows:



If we denote it by using the function $f(x)$ then the value of $f(x)$ will be $1.f(x)=1$; The intermediate node is II. Now let us consider the case 2×2 matrix the value here will be $f(x)=1+2=3$; The intermediate nodes are $1(2,3), 2(4)$.



For the case for the 3×3 matrix the value of the function $f(x)=1+2+2=5$;



Similarly for the 4×4 matrix we can get the value of $f(x)=1+2+2+2$.

Here in this case we were having only 4 elements in a ring. Suppose we have 8 elements in the ring in that case we have to compute the number of nodes required to communicate or establish the connection between the farthest nodes.

Justification - Let us consider the case of 1×1 matrix to communicate between the farthest node we need 3 nodes, i.e., $f(x)=3$. In case of 2×2 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=3+4$; In case of 3×3 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=7+8+8$; In case of 4×4 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=7+8+8+8$; Now the total number of nodes can be derived by the general formula as $(N/2-1)+(M-1)*(N/2)$ where N = number of nodes present in the unit or element, M = dimension of the square matrix. The data can be represented in the tabular form as follows:

nodes, i.e., $f(x)=3+4+4$; In case of 4×4 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=3+4+4+4$; In case of 16 elements in a ring, we can proceed as follows. Let us consider the case of 1×1 matrix to communicate between the farthest node we need 3 nodes, i.e., $f(x)=7$. In case of 2×2 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=7+8$; In case of 3×3 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=7+8+8$; In case of 4×4 matrix to communicate between the farthest node we need 7 nodes, i.e., $f(x)=7+8+8+8$; Now the total number of nodes can be derived by the general formula as $(N/2-1)+(M-1)*(N/2)$ where N = number of nodes present in the unit or element, M = dimension of the square matrix. The data can be represented in the tabular form as follows:

No. of nodes	1×1	2×2	3×3	4×4
4	1	3	5	7
8	3	7	11	15
16	7	15	23	31

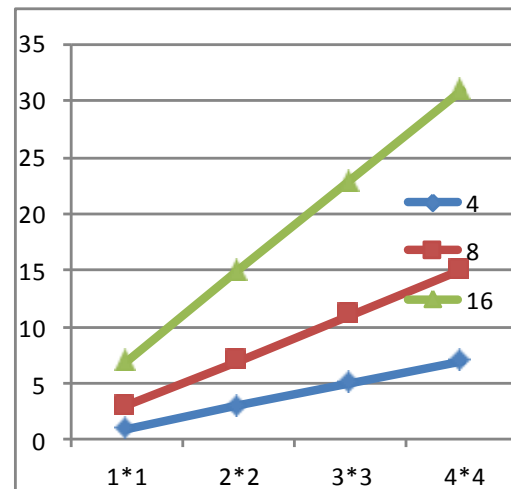


Fig. 2: Nodal communication in cluster

The x-axis represents the $M \times M$ matrix where M varies from 1 to 3. The y-axis represents the number of optimum communication nodes required in the establishing the path between the source node and the farthest node. The number of nodes per element is indicated by the 3 colors.

III. STATISTICAL INFERENCE OF FUTURISTIC VALUES

In statistical inferences the input & output of a situation are related with a certain relation or function based on which we infer futuristic values. Consider a real-time situation in which a given input parameter is observed over time between instants T_1 & T_2 given the relation [2]

$$M_t = a.e^t \text{ then } M_{avg} = \sqrt{(M_{t1} \cdot M_{t2})}$$

Case 1:

If we take observations at equal instants of time then

$$M_{t1} = a.e^{t1}$$

$$M_{t2} = a.e^{t1+k}$$

$$M_{t3} = a.e^{t1+2k}$$

General term $M_{tn} = a.e^{t1+(n-1)k}$ i.e. the values of output M forms a G.P. series of increasing order common ratio as e^k .

Case 2:

If we take observation at unequal timing interval in that case

$$T1 = t1 \Rightarrow M_{t1} = a.e^{t1}$$

$$T2 = t1 + k1 \Rightarrow M_{t2} = a.e^{t1+k1}$$

$$T3 = t2 + k2 = t1 + (k1 + k2) \Rightarrow M_{t2} = a.e^{t2+k2} = M_{t2} = a.e^{t1+k1+k2}$$

General term $Tn = T1+(k1+k2+k3+...+kn)$

$$Tn = tn-1 + kn-1 = t1 + (k1 + k2 + k3 + ... + kn-1) \Rightarrow$$

$$M_{tn} = a.e^{tn-1+kn-1} = M_{tn} = a.e^{(t1+k1+k2+k3+...+kn-1)} = a.e^{t1+Ktotal} \text{ i.e.}$$

now any futuristic value say at instant tn is

$$M_{tn} = a.e^{t1}.e^{Ktotal} \text{ (observed value)}$$

Given $M_t = a.e^t$, taking log on both sides we have,

$$\ln(M_t) = \ln(a) + t$$

$$\text{i.e. } \ln(M_{tn}) = \ln(a) + tn$$

$$\ln(M_{tn}) = \ln(a) + t1+k1+k2+k3+...+kn-1$$

Thus we have obtained a log linear model for the above function $M_t = a.e^t$ using which we can calculate or predict the futuristic values for increased ranges.

$$Y = m.X + C$$

If we try to minimize the value of Ktotal we can do so by making $k1=k2=k3=...=kn-1$ which is same as Case 1.

IV. PROJECTION OF SENSED INFORMATION

Let $I = \{i_1, i_2, \dots, i_n\}$ be the set of sensed information. In the process of feature appropriate observation, forward selection, backward elimination and decision based induction methods are applied.

A. Forward selection based information sensing

Let $I = \{i_1, i_2, \dots, i_n\}$ be the set of information estimates of various trends noted after observation in respective timing instants $Y = \{y_1, y_2, \dots, y_n\}$. The accuracy measurement is to be calculated first based on comparison analysis. The minimum deviation reflects high accuracy level of prediction and that information will be selected. In this manner, $\{ \}$, $\{ \text{best information} \}$, $\{ \text{first two} \}$will be selected.

B. Backward elimination based information sensing

Using backward elimination, in each stage each information is eliminated and thereby after the final screening stage the projected set reveals the final optimum information space.

C. Cumulative frequency based information sensing

OBSERVATIONS	INFORMATION INVOLVED
g_1	i_1, i_3, i_4, i_6
g_2	i_3, i_5
g_3	i_4, i_5, i_6
g_4	i_2, i_3, i_5
g_5	i_1, i_2
g_6	i_1, i_2, i_3, i_6

Table1 : Association of information against each observation

Features	Initial value	Count	Value	(Value) ²
i_1	0.1	3	0.3	0.09
i_2	0.2	3	0.6	0.36
i_3	0.3	4	1.2	1.44
i_4	0.4	2	0.8	0.64
i_5	0.5	3	1.5	2.25
i_6	0.6	3	1.8	3.24

Table 2 : Determination of count and value

Now CF = (x , y , z)

where x = number of elements , y = linear sum of the elements and z = sum of the square of the elements[3]

V. BINARY TREE BASED GAIN CLASSIFIER

In this section information represents gain analysis. A search[4] can be formed based on the initial search term and its gradual sub term while the process of matching. Thereby the level is increased, in initial search term is the root and the final term fully matching with the context of the users' desire is a leaf node.

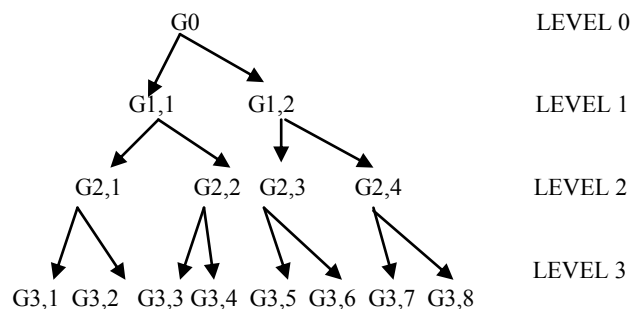


Fig3: Binary tree based gain classifier

In the above figure, G0 is the root that is initial search term. If a user wants to analyze further gain classification, then identify each search term as a binary code and by giving the code number he can analyze the position of gain estimate in the model. The concept of coding is as follows:

Value = 0 if the search term is a left child of parent node
= 1 otherwise

C3 = { 10 }
C4 = { 11 }

Theorem: In the process of coding, $\sum_{i=1}^N 1/2^{L_i} = 1$, where

L_i is the length of code of i th leaf node in the tree, N is total number of leaf nodes and $1 < i < N$.

Proof:

From fig.3 codes of leaf nodes are as follows:

Nodes	Respective code
G3,1	000
G3,2	001
G3,3	010
G3,4	011
G3,5	100
G3,6	101
G3,7	110
G3,8	111

So, $N=8$. Each leaf node has identical code length i.e. 3.
Therefore, $1/2^{L_i} = 1/2 = 1/8$, $1/2 = 1/8$, ... $1/2 = 1/8$

We now design a binary tree based classifier taking some parameters for examination purpose and represent each point on the basis of a code generated by arithmetic coding. Finally, represent the same on the basis of set theory. We assume that the gain set available is $G = \{g_1, g_2, g_3, g_4\}$. The parameters based on which the examination is to be carried are the elements of the set $P = \{p_1, p_2, p_3\}$. The result of the examination are denoted in the form of Boolean variables such that the outputs are denoted as:

NO = 0

YES = 1

At the initial timing instant, the parameter p_1 is applied for testing purpose. Hence, in the initial stage, there will be at least one class while a maximum of two classes. In the second level, the parameter p_2 is applied and accordingly the classes are defined. In the final stage, the parameters p_3 is applied.

If we assume the classifier as a binary tree representation, we can apply arithmetic coding to each class such that a 'NO' of a particular exam is denoted by '0' and a 'YES' is denoted by '1'. In the initial stage, the class which contains the elements for negative supply of p_1 is $C1 = \{d_1, d_2\}$, while, $C2 = \{d_2, d_4\}$. In this manner, the tree is to be constructed such that the code word for each class is denoted by ijk where $i \in \{0, 1\}$, $j \in \{0, 1\}$ and $k \in \{0, 1\}$.

For p_1 :

C1 = { d1,d2 }
C2 = { d2,d4 }

For p_2 :

C1 = { 00 }
C2 = { 01 }

For p_3 :

Class	ijk	False	True
C1	000	p_1, p_2, p_3	-
C2	001	p_1, p_2	p_3
C3	010	p_1, p_3	p_2
C4	011	p_1	p_2, p_3
C5	100	p_2, p_3	p_1
C6	101	p_2	p_1, p_3
C7	110	p_3	p_1, p_2
C8	111	-	p_1, p_2, p_3

In the initial stage, classes are $C1, C2$ based on the parameter p_1 . In the second stage, the classes are $C1, C2, C3$ based on p_2 . In the last stage, classes are $C1, C2, \dots, C8$ based on p_3 . This means that if we assume that 'n' is the number of parameters involved in the system for examination purpose. Then, the maximum length of code word for a particular class is 'n'. The number of classes is 2^n , provided that the classes are distinct in nature.

VI. CODED INFORMATION SENSING

Let original message is "FATHER". For the first alphabet, $\mu_{\text{value}} = 1/((\text{position of that}) + \pi/100)$. Hence it's offset value = ceiling of (the product of μ_{value} and 10). The weight is given by its position in alphabet string[5].

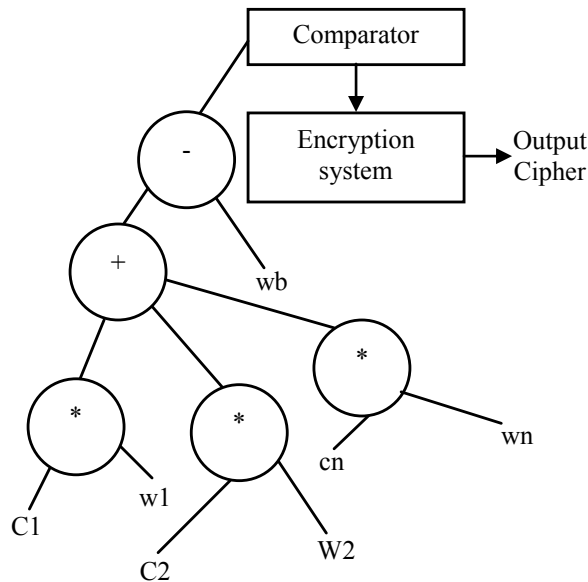
Therefore total_value = offset value * weight. From the next character onwards, $\mu_{\text{value_next}} = 1/(\text{mod value of } (\text{position of next} - \text{position of previous}) + \pi/100)$. Hence total_value is calculated in similar manner. Now, bias value will be equal to total number of characters in the message. Compute net_value as (total_value_first char + total_value_last char) - (bias_value) and let it be x (say).

Mode	Operation
$0 \leq x < 100$	Reverse the message.
$100 \leq x < 150$	Circular left shift of message by $n/2$ bits where n = bias value.
$150 \leq x < 200$	Circular right shift of message by $n/2$ bits

Iteration 1: $\mu_F = 1/((\text{position of 'F' in alphabet list}) + \pi/100) = 1/((6) + \pi/100) = 0.165798547$. Offset value = ceiling of $(0.165798547 * 10) = 2$. Weight = position of 'F' in alphabet list = 6. Thus, total_value = $2 * 6 = 12$.

Iteration 2: $\mu_A = 1/((\text{position of 'A' - position of 'F'}) + \pi/100) = 1/((1-6) + \pi/100) = 1/5.031415927 = 0.198751209$. Offset value = ceiling of $(0.198751209 * 10) = 2$. Weight = 1. Thus, total_value = $2 * 1 = 2$.

Iteration 3: $\mu_T = 1/((\text{position of 'T' - position of 'A'}) + \pi/100) = 1/((20-1) + \pi/100) = 1/19.03141593 = 0.052544697$. Offset value = ceiling of $(0.052544697 * 10) = 1$. Weight = 20. Thus, total_value = $1 * 20 = 20$.



C_i = offset value for $i = 1$ to n , w_i = weight, w_b = bias value

Fig 4: Coding Model

Iteration 4: $\mu_H = 1/(|(\text{position of 'H'} - \text{position of 'T'})| + \pi/100) = 1/(|(8-20)| + \pi/100) = 1/12.03141593 = 0.083115736$. Offset value = ceiling of $(0.083115736 * 10) = 1$. Weight = 8. Thus, $\text{total_value} = 1 * 8 = 8$.

Iteration 5: $\mu_E = 1/(|(\text{position of 'E'} - \text{position of 'H'})| + \pi/100) = 1/(|(5-8)| + \pi/100) = 1/3.031415927 = 0.32987885$. Offset value = ceiling of $(0.32987885 * 10) = 4$. Weight = 5. Thus, $\text{total_value} = 4 * 5 = 20$.

Iteration 6: $\mu_R = 1/(|(\text{position of 'R'} - \text{position of 'E'})| + \pi/100) = 1/(|(18-5)| + \pi/100) = 1/13.03141593 = 0.076737632$. Offset value = ceiling of $(0.076737632 * 10) = 1$. Weight = 18. Thus, $\text{total_value} = 1 * 18 = 18$.

Now, w_b = bias value = number of bits in FATHER = 6. So $\text{net_value} = \text{accumulated sum of all total_value} - w_b = (12 + 2 + 20 + 8 + 20 + 18) - 6 = 74$. It falls in the range $0 \leq x < 100$. So, "FATHER" is reversed.

Therefore resultant cipher is "REHTAF".

VII. CONCLUSION

The paper points out information merging in grid and cluster network models. Statistical means of information prediction as well as forward, backward and cumulative frequency based schemes have been analyzed. Binary tree based information classification and coded information have been justified with relevant mathematical analysis.

REFERENCES

- [1] A.Kumar, P.Chakrabarti, P.Saini, V.Gupta, "Proposed techniques of random walk, statistical and cluster based node realization" communicated to IEEE conf. Advances in Computer Science ACS 2010, India, Dec10
- [2] P.Chakrabarti, S.K.De, S.C.Sikdar, "Statistical Quantification of Gain Analysis in Strategic Management" published in IJCSNS, Korea, Vol 9 No11, pp.315-318, 2009
- [3] P.Chakrabarti, "Data mining- A Mathematical Realization and cryptic application using variable key" published in International journal, Advances in Information Mining, Vol 2 No 1, pp-18-22,2010
- [4] P.Chakrabarti, P.S.Goswami, "Approach towards realizing resource mining and secured information transfer" published in international journal of IJCSNS, Korea, Vol 8 No.7, pp345-350, 2008
- [5] P.Chakrabarti, "Attacking Attackers in Relevant to Information Security" Proceedings of RIMT-IET, Mandi Gobindgarh. pp 69-71, March 29, 2008

About authors:



Debasis Mukherjee (20/08/80) is pursuing Ph.D. from USIT, GGSIPU, Delhi, India from 2010. He received the M. Tech. degree in VLSI Design from CDAC Noida in 2008 and bachelor degree in Electronics and Instrumentation Engineering from BUIE, Bankura, West Bengal, India in 2003. He achieved first place in district in "Science Talent Search Test" 1991. He has some publications of repute in IEEE conferences.



Dr.P.Chakrabarti(09/03/81) is currently serving as Associate Professor in the department of Computer Science and Engineering of Sir Padampat Singhanian University, Udaipur. Previously he worked at Bengal Institute of Technology and Management, Oriental Institute of Science and Technology, Dr.B.C.Roy Engineering College, Heritage Institute of Technology, Sammilani College. He obtained his Ph.D(Engg) degree from Jadavpur University in Sep09, did M.E. in Computer Science and Engineering in 2005, Executive MBA in 2008 and B.Tech in Computer Science and Engineering in 2003. He is a **life member** of Indian Science Congress Association, Calcutta Mathematical Society, Calcutta

Statistical Association , Indian Society for Technical Education , Cryptology Research Society of India, IAENG(HongKong), CSTA(USA), **annual member** of Computer Society of India, VLSI Society of India , IEEE(USA), **senior member** of IACSIT(Singapore) and **selected member** of The IAENG Society of Artificial Intelligence , Computer Science , Data Mining. He is a Reviewer of International journal of Information Processing and Management (Elsevier) , International Journal of Computers and Applications , Canada and International Journal of Computer Science and Information Security(IJCSIS,USA), editorial board member of International Journal of Engineering and Technology, Singapore and International Journal of Computer and Electrical Engineering. He has about 100 papers in national and international journals and conferences in his credit and two patents(filed). He has several visiting assignments at BHU Varanasi , IIT Kharagpur , Amity University,Kolkata , et al.

A.Khanna and V.Gupta are the third year students of Information Technology and Computer Science & Engg. branch respectively of Sir Padampat Singhania University.

Scaling Apriori for Association Rule Mining using Mutual Information Based Entropy

S.Prakash, Research Scholar
Sasurie College of Engineering
Vijayamangalam, Erode(DT)
Tamilnadu, India. Ph.09942650818
Mail:prakash_ant2002@yahoo.co.in

Dr.R.M.S.Parvathi M.E.(CSE), Ph.D.
Principal
Sengunthar College of Engg. for Women
Tiruchengode, Tamilnadu, India
rmsparvathi@india.com

Abstract - Extracting information from large datasets is a well-studied research problem. As larger and larger data sets become available (e.g., from customer behavior data from organizations such as Wal-Mart) it is getting essential to find better ways to extract relations (inferences) from them. This thesis proposes an improved Apriori algorithm to minimize the number of candidate sets while generating association rules by evaluating quantitative information associated with each item that occurs in a transaction, which was usually, discarded as traditional association rules focus just on qualitative correlations. The proposed approach reduces not only the number of item sets generated but also the overall execution time of the algorithm. Any valued attribute will be treated as quantitative and will be used to derive the quantitative association rules which usually increases the rules' information content. Transaction reduction is achieved by discarding the transactions that does not contain any frequent item set in subsequent scans which in turn reduces overall execution time. Dynamic item set counting is done by adding new candidate item sets only when all of their subsets are estimated to be frequent. The frequent item ranges are the basis for generating higher order item ranges using Apriori algorithm. During each iteration of the algorithm, use the frequent sets from the previous iteration to generate the candidate sets and check whether their support is above the threshold. The set of candidate sets found is pruned by a strategy that discards sets which contain infrequent subsets. The thesis evaluate the scalability of the algorithm by considering transaction time, number of item sets used in the transaction and memory utilization. Quantitative association rules can be used in several domains where the traditional approach is employed. The unique requirement for such use is to have a semantic connection between the components of the item-value pairs. The proposal used mutual information based on entropy to generate association rules from non-biological datasets.

Keywords-Apriori, Quantitative attribute, Entropy

I. INTRODUCTION

Data mining, also known as knowledge discovery in databases, has been recognized as a new area for dataset research. The problem of discovering association rules was introduced in latter stages. Given a set of transactions, where each transaction is a set of items, an association rule is an expression of the form $X \rightarrow Y$, where X and Y are sets of items. The problem is to find all association rules that satisfy user-specified minimum support and minimum confidence constraints.

Conceptually, this problem can be viewed as finding associations between the "1" values in a relational table where all the attributes are Boolean. The table has an attribute corresponding to each item and a record corresponding to each transaction. The value of an attribute for a given record is "1" if the item corresponding to the attribute is present in the transaction corresponding to the record, "0" else.

Relational tables in most business and scientific domains have richer attribute types. Attributes can be quantitative (e.g. age, income) or categorical (e.g. zip code, make of car). Boolean attributes can be considered a special case of categorical attributes. This thesis defines the problem of mining association rules over quantitative attribute in large relational tables and present techniques for discovering such rules. This is referred as the Quantitative Association Rules problem.

The problem of mining association rules in categorical data presented in customer transactions was introduced by Agrawal, Imielinski and Swami [1][2]. This thesis work provided basic idea to several investigation efforts [4] resulting in descriptions of how to extend the original concepts and how to increase the performance of the related algorithms.

The original problem of mining association rules was formulated as how to find rules of the form $set1 \rightarrow set2$. This rule is supposed to denote affinity or correlation among the two sets containing nominal or ordinal data items. More specifically, such an association rule should translate the following meaning: customers that buy the products in *set1* also buy the products in *set2*. Statistical basis is represented in the form of minimum support and confidence measures of these rules with respect to the set of customer transactions.

The original problem as proposed by Agrawal et al.[2] was extended in several directions such as adding or replacing the confidence and support by other measures, or filtering the rules during or after generation, or including quantitative attributes. Srikant and Agrawal describe a new approach where quantitative data can be treated as categorical. This is very important since otherwise part of the customer transaction information is discarded.

Whenever an extension is proposed it must be checked in terms of its performance. The algorithm efficiency is linked to the size of the dataset that is amenable to be treated. Therefore it is crucial to have efficient algorithms that enable us to examine and extract valuable decision-making information in the ever larger databases.

This thesis presents an algorithm that can be used in the context of several of the extensions provided in the literature but at the same time preserves its performance. The approach in our algorithm is to explore multidimensional properties of the data (provided such properties are present), allowing to combine this additional information in a very efficient pruning phase. This results in a very flexible and efficient algorithm that was used with success in several experiments using quantitative databases with performance measure done on the memory utilization during the transactional pruning of the record sets.

II. LITERATURE REVIEW

Various proposals for mining association rules from transaction data were presented on different contexts. Some of these proposals are constraint-based in the sense that all rules must fulfill a predefined set of conditions, such as support and confidence [6,7,8]. The second class identifies just the most interesting rules (or optimal) in accordance to some interestingness metric, including confidence,

support, gain, chi-squared value, gini, entropy gain, laplace, lift, and conviction [9,6]. However, the main goal common to all of these algorithms is to reduce the number of generated rules.

A) Existing Scheme

The thesis extends the first group of techniques since it does not relax any set of conditions nor employ an interestingness criteria to sort the generated rules. In this context, many algorithms for efficient generation of frequent item sets have been proposed in the literature since the problem was first introduced in [10]. The DHP algorithm [11] uses a hash table in pass k to perform efficient pruning of $(k+1)$ -item sets. The Partition algorithm minimizes I/O by scanning the dataset only twice. In the first pass it generates the set of all potentially frequent item sets, and in the second pass the support for all these is measured. The above algorithms are all specialized techniques which do not use any dataset operations. Algorithms using only general purpose DBMS systems and relational algebra operations have also been proposed [9,10].

Few other works try to solve this mining problem for quantitative attributes. In [5], the authors proposed an algorithm which is an adaptation of the Apriori algorithm for quantitative attributes. It partitions each quantitative attribute into consecutive intervals using *equi-depth* bins. Then adjacent intervals may be combined to form new intervals in a controlled manner. From these intervals, *frequent item sets* (c.f. *large item sets* in Apriori Algorithm) will then be identified.

Association rules will be generated accordingly. The problems with this approach is that the number of possible interval combinations grows exponentially as the number of quantitative attributes increases, so it is not easy to extend the algorithm to higher dimensional cases. Besides, the set of rules generated may consist of redundant rules for which they present a “greater-than-expected-value” interest measure to identify the interesting ones.

Some other efforts that exploit quantitative information present in transactions for generating association rules [12]. In [5], the quantitative rules are generated by discrediting the occurrence values of an attribute in fixed-length intervals and applying the standard Apriori algorithm for generating association rules. However, although simple, the rules generated by this approach may not be intuitive, mainly when there are semantic intervals that do not match the partition employed.

Other authors [5] proposed novel solutions that minimize this problem by considering the distance among item quantities for delimiting the intervals, that is, their “physical” placement, but not the frequency of occurrence as a relevance metric.

To visualize the information in the massive tables of quantitative measurements we plan to use clustering and mutual information based on entropy. Clustering is an old studied technique used to extract this information from customer behavior data sets. This follows from the fact that related customer purchase through word of mouth have similar patterns of customer behavior. Clustering groups records that are “similar” in the same group. It suffers from two major defects. It does not tell you how the two customer buyin behavior/clusters are exactly related. Moreover, it gives you a global picture and any relation at a local level can be lost.

B) Proposed Scheme

The proposed scheme comprises of two phases. The first phase of the thesis concerns about the quantitative association rule mining with the enhancement on Apriori algorithm. The second phase of the thesis deals with the reduction of memory utilization during the pruning phase of the transactional execution.

The algorithm for generating quantitative association rules starts by counting the item ranges in the data set, in order to determine the frequent ones. These frequent item ranges are the basis for generating higher order item ranges using an algorithm similar to Apriori. Take into account the size of a transaction as the number of items that it comprises.

- a) Define an item set m as a set of items of size m
- b) Specify frequent (large) item sets by F_m
- c) Specify candidate item sets (possibly frequent) by L_m .

A n range set is a set of n - item ranges, and each m -item set has a n -range set that stores the quantitative rules of the item set. During each iteration of the algorithm, the system uses the frequent sets from the previous iteration to generate the candidate sets and check whether their support is above the threshold. The set of candidate sets found is pruned by a strategy that discards sets which contain infrequent subsets. The algorithm ends when there are no more candidates' sets to be verified.

The enhancement of Apriori is done by increasing the efficiency of candidate pruning phase by reducing the number of candidates that are generated to further verification. The proposed algorithm use quantitative information to estimate more precisely the overlap in terms of transactions. The basic elements considered in the development of the algorithm are number of transactions, average size of transaction, average size of the maximal large item sets, number of items, and distribution of occurrences of large item sets.

The second phase of the thesis claimed improvement over A priori by considering memory consumption for data transaction. This part of the algorithm generate all candidates based on 2-frequent item sets on sorted dataset and already generates all frequent item sets that can no longer be supported by transactions that still have to be processed. Thus the new algorithm no longer has to maintain the covers of all past item sets into main memory. In this way, The proposed level-wise algorithms accesses a dataset less often than Apriori and require less memory because of the utilization of additional upward closure properties.

C) Mutual Information based entropy

The mutual information $I(X, Y)$ measures how much (on average) the realization of random variable Y tells us about the realization of X , i.e., how by how much the entropy of X is reduced if we know the realization of Y .

$$I(X; Y) = H(X) - H(X|Y)$$

For example, the mutual information between a cue and the environment indicates us how much on average the cue tells us about the environment. The mutual information between a spike train and a sensory input tells us how much the spike train tells us about the sensory input. If the cue is perfectly informative, if it tells us everything about the environment and nothing extra, then the mutual information between cue and environment is simply the entropy of the environment:

$$I(X; Y) = H(X) - H(X|Y) = H(X) - H(X|X) = H(X).$$

In other words, the mutual information between a random variable and itself is simply its entropy: $I(X; X) = H(X)$. Surprisingly, mutual information is symmetric; X tells us exactly as much about Y as Y tells us about X .

III. QUANTITATIVE ASSOCIATION RULE MINING – MUTUAL INFORMATION BASED ENTROPY

The proposal of this work use mutual information based on entropy for generating quantitative association rules. Apart from the usual positive correlations between the customers, this criterion would also discover association rules with negative correlations in the data sets. It is expected to find results of the form attrib 1/ attrib 2 \rightarrow ^attrib3, which can be interpreted as follows: Attrib1 and Attrib2 are co expressed and have silencing effect on attrib 3. Then compare the results from our experiments to those obtained from clustering.

First tune various parameters (like support, support fraction, significance level), of the auto performance dataset. This was because even with binary data, 2468 attributes may lead to the power(2, 2468) relations (which the software was not designed to handle). Here, it is needed to know that for the problem under consideration, the auto are attributes, as needed to find relationships among them. To overcome this problem used another approach, in which data attributes were already known to be related, using the results obtained from clustering. This decreases the number of attributes to manageable levels (both for program). The proposed work used the approach above to find the relationships (positive, negative) among the attributes.

Algorithm Steps

- a. Find all frequent item sets (i.e., satisfy minimum support)
- b. Generate strong association rules from the frequent item sets (each rule must satisfy minimum support and minimum confidence).
- c. Identify the quantitative elements
- d. Sorting the item sets based on the frequency and quantitative elements
- e. Merge the more associated rules of item pairs
- f. Discard the infrequent item value pairs
- g. Iterate the steps c to f till the required mining results are achieved
- h.

Let $I = \{i_1, i_2 \dots i_m\}$ be a set of items, and T a set of transactions, each a subset of I . An association rule is an implication of the form $A \Rightarrow B$, where A and B are non-intersecting. The support of $A \Rightarrow B$ is the percentage of the transactions that contain both A and B . The confidence of $A \Rightarrow B$ is the percentage of transactions containing A that also

contain B (interpret as $P(B|A)$). The occurrence frequency of an item set is the number of transactions that contain the item set.

IV. IMPLEMENTATION OF QUANTITATIVE APRIORI

The function `op` is an associative and commutative function. Thus, the iterations of the `foreach` loop can be performed in any order. The data-structure `Reduc` is referred to as the reduction object. The main correctness challenge in parallelizing a attribute like this on a shared memory machine arises because of possible race conditions when multiple processors update the same element of the reduction object.

The element of the reduction object that is updated in a loop iteration is determined only as a result of the processing. In the a priori association mining algorithm, the data item read needs to be matched against all candidates to determine the set of candidates whose counts will be incremented. The major factors that make these loops challenging to execute efficiently and correctly are as follows:

It is not possible to statically partition the reduction object so that different processors update disjoint portions of the collection. Thus, race conditions must be avoided at runtime.

The execution time of the function process can be a significant part of the execution time of an iteration of the loop. Thus, runtime preprocessing or scheduling techniques cannot be applied.

The updates to the reduction object are fine grained. The reduction object comprises a large number of elements that take only a few bytes, and the for each loop comprises a large number of iterations, each of which may take only a small number of cycles.

```
{ * Outer Sequential Loop *}
While() {
  { * Reduction Loop*}
  Foreach(element e ) {
    ( i, val) = process (e);
    Reduc(i) = Reduc(i) op val;
  }
}
```

Fig 1: Pseudo code

The consumer behavior auto databases obtained data from UCI Machine Learning Repository. The data obtained was about CPU-performance and automobile mileage. The data was discretized into binary values. For these data sets the

discretization was done in accordance with interpretation required. This discretization was done automatically using the written software. This software also formatted the data into the format required by the program. A finer level of discretization (or supporting the real values) would have been more appropriate, but the used approach also gave much of the useful results.

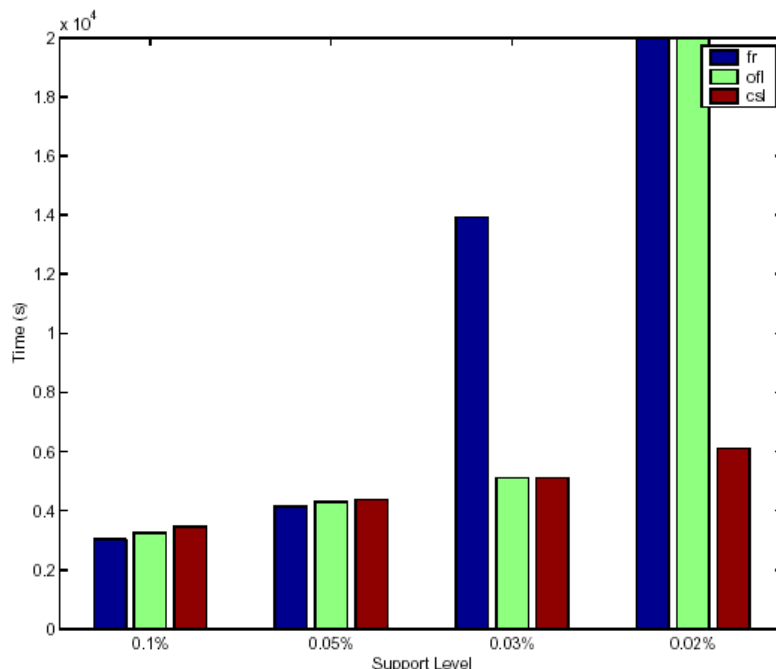
V. EXPERIMENTAL RESULTS FROM APRIORI

The process of executing quantitative Association rule mining for the auto manufacturer power evaluation data is given below

- Get data (file: auto-mpg.data): 9 attributes, 398 samples.
- Remove unique attributes (IDs). Here, car_name attribute has been removed.
- Remove those samples (total 5) that contain “?” (missing data) as a value for some of their attributes (so, we are left with 8 attributes and 393 samples).
- Discretize real-valued attributes based on their average values (which is (maximum attribute value + minimum attribute value) / 2)
- Run the program to generate association rules using mutual information based on entropy metric.

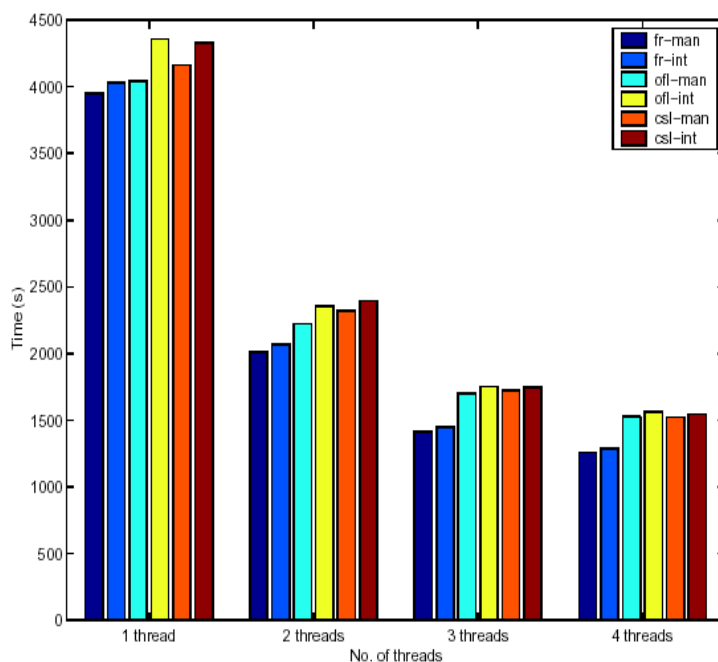
The experiment focused on evaluating all quantitative a priori techniques. Since we were interested in seeing the best performance, we used banking data set samples. We used a 1 GB dataset. A confidence of 90% and support of 0.5 is used.

Execution times using 1, 2, 3, and 4 threads are presented on the processor. With 1 thread, Apriori does not have any significant overheads as compared to the sequential version. Therefore, this version is used for reporting all speedups. Though the performance of quantitative a priori is considerably lower than a priori, they are promising for the cases when sufficient memory for supporting full replication may not be available. We consider four support levels, 0.1%, 0.05%, 0.03%, and 0.02%. The execution time efficiency is improved for the quantitative a priori on frequent item set evaluation with the support count (Graph 1).



Graph 1: Support VS Time on Quantitative and qualitative A priori

The thread execution on the quantitative a priori and qualitative a priori are evaluated for the same data set (Graph 2). Here the initial thread requires more time, however consequent threads shows better scalable performance of quantitative Apriori.



Graph 2: Thread Vs Time for a priori execution on rule set generation

By observing the output from the program it is seen that a few relationships between the attributes had high values of mutual information. Namely, the highest MI-values were obtained for:

- a) displacement and horsepower. Further, by observing the entropy values we may notice that there are very few cars that have small displacement and high horsepower.
- b) displacement and weight. Further, by observing the entropy values we may notice that there are very few cars that have large displacement and light weight.
- c) cylinders and weight. Further, by observing the entropy values we may notice that there are very few cars that have small number of cylinders and heavy weight.
- d) horsepower and weight. Further, by observing the entropy values we may notice that there are very few cars that have large horsepower but heavy weight.

VI. CONCLUSION

The thesis have defined a new a rule set namely the informative rule set that presents prediction sequences equal to those presented by the association rule set using the confidence priority. The informative rule set is significantly smaller than the association rule set, be especially when the minimum support is small.

The proposed method has some merit in extracting information from huge data sets by pruning the initial information (to bring it down to the manageable levels) and then finding the association rules among the attributes. Further, the approach is used to predict the relationships among the silencer auto power, weight, model, year etc., could be extended to unknown function.

The proposed scheme have characterized the relationships between the informative rule set and the non-redundant association rule set, and revealed that the informative rule set is a subset of the non-redundant association rule set. The thesis considers the upward closure properties of informative rule set for omission of uninformative association rules, and presented a direct algorithm to efficiently generate the informative rule set without generating all frequent item sets.

The informative rule set generated in this thesis is significantly smaller than both the association rule set and the non-redundant association rule set for a given dataset that can be generated more efficiently than the association rule set. The

efficiency improvement results from that the generation of the informative rule set needs fewer candidates and dataset accesses than that of the association rule set rather than large memory usage like some other efficient algorithms.

REFERENCES

- [1] R. Agrawal, T. Imielinski, and A. Swami. Dataset mining: A performance perspective. In *IEEE Transactions on Knowledge and Data Engineering*, December 1993.
- [2] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD* Washington, D.C., pages 207-216, May 1993.
- [3] R. Miller and Y. Yang. Association rules over interval data. In *ACM SIGMOD Conference*, Tucson, Arizona, pages 452 - 461, May 1997.
- [4] J. Park, M. Chen, and P. Yu. An effective hash based algorithm for mining associative rules. In *ACM SIGMOD Conference*, San Jose, CA, pages 175 - 186, May 1995.
- [5] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. In *Proceedings of the ACM SIGMOD Conference on Management of Data*, pages 1-12, Montreal, Canada, June, 1996.
- [6] R. Agrawal, T. Imielinski, and A. Swami. Dataset mining: A performance perspective. In *IEEE Transactions on Knowledge and Data Engineering*, December 1993.
- [7] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo. Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, San Jose, CA, pages 307-328, 1996.
- [8] R. Bayardo, R. Agrawal, and D. Gunopulos. Constraint- based rule mining in large, dense databases. In *15th International Conference on Data Engineering*, Sydney, Australia, pages 188 - 197, March 1999.
- [9] R. Bayardo and R. Agrawal. Mining the most interesting rules. In *5th ACM SIGKDD International Conference on Knowledge*, San Diego, CA, Pages 145 - 154, August 1999.
- [10] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD* Washington, D.C., pages 207-216, May 1993.
- [11] J. Park, M. Chen, and P. Yu. An effective hash based algorithm for mining associative rules. In *ACM SIGMOD Conference*, San Jose, CA, pages 175 - 186, May 1995.
- [12] Prakash.S and R.M.S.Parvathi. " Scaling Apriori for Association Rule Mining Efficiency ".*Proceedings of the Fourth International Conference*, Amrutvani College of Engineering, Sangamner,Maharatra.pp 29, March 2009.

AUTHORS PROFILE



Prof. S. Prakash has completed his M.E., (Computer Science and Engineering) in K.S.R.College of Technology , Tamilnadu, India in 2006. Now he is doing research in the field of Association Rule Mining algorithms. Currently, he is working as Assistant Professor in the department of Information Technology, Sasurie College of Engineering, and Tamilnadu. India. He has completed 9 years of teaching service.



Dr. R.M.S. Parvathi has completed her Ph.D., degree in Computer Science and Engineering in 2005 in Bharathiar University, Tamilnadu, India. Currently she is a Principal and Professor , Department of CSE in Sengunthar College of Engineering for Women, Tamilnadu, India, She has completed 20 years of teaching service. She has published more than 28 articles in International / National Journals. She has authorized 3 books with reputed publishers. She is guiding 20 Research scholars. Her research areas of interest are Software Engineering, Data Mining, Knowledge Engineering, and Object Oriented System Design.

Clustering of High-Volume Data Streams In Network Traffic

M.Vijayakumar

Department of Computer Science and Engineering ,
Sasurie College of Engineering,
Vijayamangalam, Erode(Dt), Tamilnadu, India.
tovijayakumar@gmail.com

Dr.R.M.S.Parvathi M.E.(CSE),Ph.D.

Principal,
Sengunthar College of Engineering. for Women,
Tiruchengode,Tamilnadu, India.
rmsparvathi@india.com

Abstract— The thesis concerned with the problem of mining network traffic data discovering useful associations, relationships, and groupings in large collections of data. Mathematical transformation algorithms have proven effective at reducing the content of multilingual, unstructured data into a vector that describes the content. Such methods are particularly desirable in fields undergoing information explosions, such as network traffic analysis, bio-informatics, and the intelligence community. In response, traffic mining methodology is being extended to improve performance and sufficiently scalable. The usage of data flow collected from site routers for various analysis i.e., network performance characterization, investigating computer security incidents and their prevention, network traffic statistics, and others. Currently, the data flow analysis is built as a distributed system to collect data from multiple routers, both at the edge of the site network as well as from local routers and multilayer switches. Average per day volume is about 2GBytes of raw data. Despite a high volume of collected information, some analysis is conducted in near real time to satisfy demands of users communities for quick results. The proposed work present an efficient clustering means to analyze experimental results for traffic data streams nature (symmetric and asymmetric). As summary, this paper describe a system designed to satisfy three primary goals i.e., real-time concept mining of high-volume data streams, dynamic data flow into a relational hierarchy; and adaptive reorganization of the traffic data hierarchy in response to evolving circumstances and network traffic time to time. The proposed clustering network traffic data flow collection and analysis system describe traffic characterization and network performance estimation for the data flow centre. The system checks the traffic consistency for End To End circuits and Policy Based Routing and finally, profiling of host's traffic to keep track of their typical behavior to prevent accidental blocking by site IDS system.

Keywords— Traffic analysis, network management , clustering, frequent Item set , hierarchical clustering.

I. INTRODUCTION

As part of an ongoing research project, developed a novel algorithmic approach for extracting traffic data from voluminous data streams[1][14]. The approach is applicable to internet data servers, which can be automatically identified and converted into a common structure.

Here, report on an extension of the traffic data clustering work, which clusters data flow hierarchically based on the nature requested by users, (demand, load). The new method represents the streaming hierarchical clustering of documents[2][18], which can be seen as a subfield of the nascent discipline of “streaming AI”, “evolutionary clustering”, or “AI in hardware”.

The proposed system is a High Speed Content classification system that works in three stages to classify flows of TCP traffic. A TCP flow is half of a TCP conversation. A connection from a client to a Mail server with SMTP (simple mail transfer protocol) is an example of a flow[4][26]. The connection from the Mail server back to the client is considered a separate flow. It extracts words then builds a vector representation and then scores against known nature of the data flow. The scores of the completed flows are passed out of the system for evaluation.

The base data list from stage one used for counting in stage two. Each dimension is represented by 4bits[3]. Counting for a dimension saturates at 15. When the flow ends the count vector representing the flow is passed on to stage three. In the third stage a vector representing a flow is scored against vectors representing known traffic data[10][24][25]. The data vectors are called the Score Table (ST) and are reconfigurable at run time like the WMT. The ST is derived from a set of documents. The output of the

system is a set of scores and the count array of the flow.

Evaluation of the scores determines the classification of the flow against the known traffic data. However, simply classifying the flow as the data with greatest score is not adequate. A forced classification of all flows will be undesirable in most applications. A threshold provides a confidence level to the classification of flows. Any traffic data that is not classified is considered unknown to the system. Clustering these unclassified flows is the focus of the Streaming Clustering work in this thesis.

II. HIERARCHICAL PARTITIONING

The notion that natural categories tend to be organized hierarchically as a general principle dates back at least to Aristotle, and has been refined by Simon[5][11][12]. Considering that libraries, governments, Internet newsgroups and taxonomies, and the human neocortex all organize and process information hierarchically, it is certainly a natural methodology for organizing unknown content. A logical assumption for document clustering is that like documents will group together. Groupings with large number of items are more general.

As the number of items starts to dwindle, the topic of a grouping becomes more specific. For example, the topic of "cars" is very general, while "Cadillac" is more specific and would be part of the general topic of "cars." Two standard approaches can be taken to the problem organizing a collection of documents represented as fixed length vectors of high dimensionality hierarchically agglomerative (bottom-up), and divisive (top-down)[6][13][22].

Agglomerative-Cluster Algorithm

Assign each document to a unique cluster

While more than one cluster exists

Merge the two closest clusters

Return the single remaining cluster as the root

Divisive-Cluster Algorithm

If the collection contains a single traffic data

Return it in a single cluster

Divide the collection into two parts

Call divisive-cluster on each part

The final result for both procedures is a binary tree containing a single data in every leaf, where close leaves (measured by tree-traversal distance) are related. Both methods have their general advantages and drawbacks, but for the specific problem of data

clustering where traffic data are represented via a flow sequence approach, empirical studies have shown that the top-down divisive approach produces consistently superior results.

The generally accepted explanation for these results is that local neighborhood information (which pairs of documents contain most similar word distributions) is often misleading, deceiving agglomerative clustering into making bad merge decisions early on. Divisive clustering can often avoid these mistakes by first considering the global statistics of collections of traffic data, which are more robust. Thus, a divisive hierarchical clustering approach is expected to give us the best results; the remaining decision that must be made is how to divide a collection of documents.

The centroid gives an indication of the overall makeup of a cluster, and is used in many clustering algorithms. Assuming that our set of documents, V , represents some cohesive grouping (e.g., a collection of postings from a single internet newsgroup), its centroid provides us with an indication of what concepts the "average" document refers to the value in some dimension will be between zero and one, and denote the probability of a random document from the collection scoring a hit in that dimension. It is natural to consider measuring the affinity of a traffic data to a given collection by comparing its distribution of hits to our expectations normalized by data flow vector magnitude, as data with higher magnitudes have more opportunities for hits.

The heuristic approach that is taken to attempt this is quite simple. A set of data-vectors is randomly partitioned (a fair coin is flipped for each document to assign it to either cluster one or cluster two). Data are iteratively transferred between the two clusters, one at a time, as long as doing so strictly increases the division quality. Requiring a strict increase in division quality ensures that the partitioning procedure is guaranteed to terminate[7][22].

The system will compare hierarchical partitioning to flat and hierarchical variants of the popular k-means clustering algorithm[8], and to a baseline hierarchical clustering algorithm (bisection k-means) from the synthetic dataset. These traffic data sets consist largely of flow of data at the network points, demand, and load etc.,

III. K-MEANS CLUSTERING

The k-means clustering algorithm separates input data into K groups. The number of groups, or K , is set prior to running the clustering algorithm. Each document in the data is assigned to a cluster. These assignments are used to calculate the cluster centroid or center[9][21][23]. The cluster centroids are then

utilized to determine the distance between each centroid and a data element. The algorithm seeks to minimize the inner cluster distance (i.e. form tight groups of similar data) and minimize the inter cluster distance (i.e. the groupings are non-overlapping).

The distance calculation can be performed in any number of ways. Three common methods of distance calculation include the Minkowski, Manhattan, and Euclidean Distance metrics. The cosine theta distance has also been used with k-means in order to cluster high dimensional data. The algorithm is a cyclical algorithm that performs in the following manner

- a) Initially assign document in the data to K groups
- b) Calculate the cluster centroids based on assignments
- c) For each document in the data
 - i) Recalculate distances from document to all centroids and find closest centroid
 - ii) Change document assignment to closest centroid and update the centroids that the document used to reside and currently resides
- d) Repeat step 3 until either no changes are made to document assignments or the epoch limit is reached.

Bisection k-means is a variant of the k-means algorithm. It starts with a single cluster and continually selects a cluster to split into 2 sub-clusters until the requested number of clusters is achieved.

- a) Pick a cluster to split.
- b) From the selected cluster, use k-means to cluster the elements into 2 sub clusters.
- c) Repeat steps 1 and 2 until the desired number of clusters is found.

The selection of the cluster to bisect can be done in a number of ways. For choosing the cluster with the most elements was sufficient to find good clustering, use the same heuristic here. In order to objectively compare the results of a hierarchical clustering algorithm to a flat clustering, need a means of automatically flattening a full binary cluster tree to a set of k clusters (given some particular k). A simple heuristic for doing so is to choose the (non-overlapping) subtrees that make the highest quality clusters, as defined in the previous section. So for k=2, will choose the left and right subtrees of the root. For k=3 we will take the lowest scoring of our current clusters (that is not already a leaf) and expand it into two clusters by replacing it with its two children (recall

that divisive hierarchical clustering always produces a binary tree). For bisection k-means, stop tree creation after k leaves have been formed and take them as our clusters.

The ground-truth newsgroups (horizontal axis) are ordered so that the far right column of the confusion matrix is the chaff, and the ordering of the remaining newsgroups is arbitrary. The clusters (vertical axis) are ordered based by their most frequent newsgroup, with color showing purity (blue is lowest, red highest). A perfect clustering would hence be denoted by a crisp diagonal red line. K-means is clearly inferior to the two hierarchical approaches, placing the majority of the documents into two large clusters (the two horizontal red lines near the bottom of the plot). Bisection k-means and hierarchical partitioning produce comparable results; however half of the clusters created by bisection k-means are dominated by chaff, whereas hierarchical partitioning creates only ten such "junk" clusters. This is preferable from a human analyst's point of view, as it allows uninteresting document sets to be identified and discarded more quickly.

IV. EXPERIMENTAL RESULTS

This part of the thesis describes experimental results with the streaming hierarchical partitioning algorithm. Results are presented on ISP data, streamed according to a regime designed to simulate nature of the traffic data are randomly collected, but to begin, only data from half of the ISP server are evaluated. At uniform intervals, a ISP is gradually introduced into the distribution (and hence the old newsgroup density gradually reduced).

The traffic flow appears uniformly throughout the entire data stream. In order to run streaming hierarchical partitioning clustering, need to set two parameters, the maximal number of traffic data that are capable of collecting at a time, and how often the set of data flowing in the server working memory will be reclustered. In actual deployment of course, will want to store as many traffic data as possible in fast memory, and recluster as often as possible, and the bandwidth of the data stream. Given that our dataset contains about 10GB of data should prove illustrative.

To analyze the quality of hierarchical streaming clustering, there are two basic factors consider, traffic quality as in the non-streaming case, how meaningful are the clustered data discovered by the algorithm, and traffic discovery as drift occurs, does the algorithm effectively identify new flow of traffic data.

In a controlled experiment, where know the ground-truth labeling of the data, traffic quality can be measured by considering how many non-chaff data are

assigned to clusters that are nearly “pure” (at least 90% of the documents originating in a single newsgroup). Furthermore, only consider clusters with more than 10 documents in computing this measure, henceforth referred to as a purity score.

Similarly, traffic discovery over time can be measured by considering, at any given time, how many pure clusters have been created corresponding to unique non-chaff labels, up until this time. The measure is hence cumulative, and will henceforth be referred to as a discovery score. In order to understand how effective the data insertion and removal heuristics are at augmenting hierarchical partitioning for streaming clustering. In this procedure, full hierarchical partitioning clustering will still be carried out at the same regular intervals. However, when new data are congested, traffic will be chosen entirely at random to be removed. This methodology will henceforth be referred to as naïve streaming clustering.

In order to provide a fair comparison between naïve and non-naïve streaming clustering, scores are only computed immediately after batch clustering has been completed. That is, naïve streaming clustering is not penalized for having poor results in between batch clustering. Both methods are equally effective in terms of purity scores as expected, since this score is essentially determined by the batch clustering, which is identical.

V. PERFORMANCE EVALUATION

Analysis of flow data gathered from multiple routers is the source of valuable information for various tasks. We distinguish three major areas where that information is used. These areas are computer security, performance analysis for data movement applications and verification of traffic consistence across site network infrastructure. We have considerable experience to use flow data in two first areas. The last one, verification of traffic consistency is relatively new to us. It is motivated by recent deployment End To End circuits for LHC/CMS experiment. These circuits have predictable performance characteristics and dynamic capabilities to use alternative WAN path available for the data centre.

Traffic is analyzed for different time intervals, typically 1min, 5min, 15min, 1hour, or 1 day. A start date and time can be also specified to generate historical results. Tagging of raw data can be done at any time, but for most applications, it runs when new flow sets are created and become available on NAS server. New intervals can be added as needed. The next level is the target, which represents defined list of IP blocks assigned to particular users group or community. The system Tier 1 LAN is an example of a user group address block.

	Netflow Data		Actual Traffic	
	Daily	Monthly	Daily	Monthly
Border	600MB	17GB	15TB(43TB)	300TB (600TB)
StarLight	200MB	5GB	80TB(120TB)	1.5PB(2.4PB)
CMS	1.2GB	30GB	80TB(120TB)	1.5PB(2.4PB)
CORE	3GB	40GB	N/A	N/A

Table 1: Volumes of flow data and actual traffic

A target's traffic is going to be inspected based on set of predefined filters. For example, analyze appropriate traffic to specific destinations. For each target there may be several filters defined. Filters have unique names and are stored in a database with time stamp allowing to track changes. While inspecting traffic, we may use some specific names for destinations or sources. For example, we don't need to know all details of traffic to each node of the destination cluster. Instead, we need to know traffic to whole cluster, and use a predefined name for the address block. The next level is the static definition of rules specified in separated XML files. This level is identified by name of XML configuration file. The content of configuration files may be changed over time, due different reasons, i.e. readdressing of nodes, moving them to different subnets and so on. To track these changes, the content of configuration files is stored in database with time stamp. Anything that is not statically defined will be resolved based on DNS level, i.e. top level, second, third and so on, as needed.

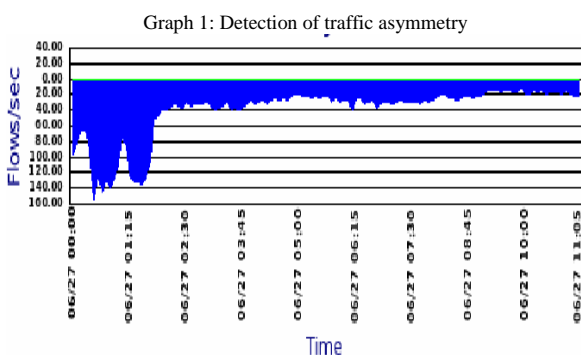
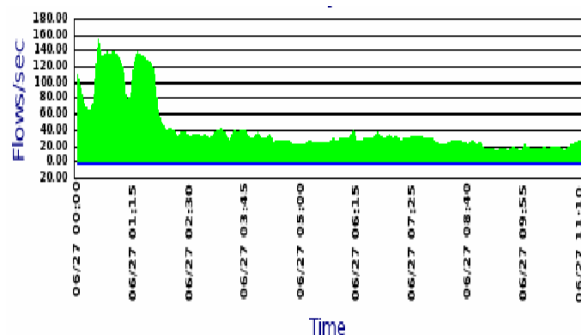
ipaddr	name	hostCount	octets	flows	packets	duration
192.18.43.225	pca-qa-fw-1.Sun.COM	53677	5523542	67229	125026	84993
221.215.83.230		39694	1834658	39695	39948	84995
83.117.22.127	c53751671.cable.wanadoo.nl	32763	3356159	32789	55019	84995
212.64.105.18	cd4406912.cable.wanadoo.nl	32762	3455589	32792	56649	84995
66.140.105.98	adsl-66-140-105-98.dsl.lbcktx.swbell.net	32757	3602024	33433	53233	84993
82.90.176.81	host81-176-static.90-82-b.business.telecomitalia.it	32750	3535797	33223	56151	84995
90.1.63.17	ALille-153-1-100-17.w90-1.abo.wanadoo.fr	32735	3091265	33065	50815	84995
208.157.165.93	ame-bb-dsl-348.dsl.airstreamcomm.net	32733	6234200	62393	102200	84995
201.66.29.248	201-66-29-248.ctaj700.dsl.brasiltelecom.net.br	32728	3870511	32751	63451	84993
83.28.73.175	lbg175.neoplus.adsl.tpnet.pl	32724	3564020	33161	56544	84993
216.171.177.46	ccc-resnet-216-171-177-46.177.171.216.in-addr.arpa	32715	2981803	33021	49002	84993

Table 1: ISP Data traffic flow output

Knowing transfer rates for data movement applications is very important. Often, neither SNMP monitoring of router/switch interfaces, nor applications themselves can provide that information. For this purpose, use passive monitoring based on net flow information.

First, on the regular basis, we generate reports with topN senders, receivers and conversations for specified interval of time between the entities that we need to monitor. The reports are produced using the results of traffic tagging described earlier. The output of the WEB based interface is very similar to Top Scanning reports.

Our objective is to detect conditions automatically and notify if it becomes a long term situation. After preliminary investigation, determined that flow data collected from routers at the edge of the campus network could be used for that purpose by comparing flow rates at various points of network. Flow rates for a symmetric path should be very close in both directions, inbound and outbound. The proposed cluster tool is used interactively to show asymmetric traffic conditions. Currently, we are working on automated alerts.



The graph 1 and 2 visualize the idea of steering production traffic. The traffic switched into a high impact path is almost always symmetric in all routers. However, traffic is going to be asymmetric at the border and the point of presence. From graphs at the right we can see that

traffic at work group router is almost symmetric. However, at the border router observed only outbound flows (second graph) and the high impact infrastructure providing alternative path, only inbound flows are observed. Thus, by comparing flow rates at the different routers along potential traffic path successfully detect asymmetry conditions.

VI. CONCLUSION

The proposed model describe a system for extracting traffic data based on the nature requested by the network administrator with an effective clustering system. The implemented system, streaming hierarchical partitioning, hierarchically clusters data streaming content.

The performance predictions include the quality of clustering, and traffic data discovery. The streaming hierarchical clustering algorithm was able to improve the ability to discover traffic data of required nature. The system has been prototyped and tested on a Xeon processor as well as on a PowerPC. To implement additional streaming functionality, some of the same circuitry can be reused in particular the computation of similarity of traffic data insertion.

In the future, plan to additionally move towards a system that, Integrates clustering into our classification system, continually searches for new and emerging traffic data sets on larger inter networks, allows the resolution of data to fade over time to allow for streaming with infinite flow of data sets.

REFERENCES

- [1] Abdun Naser Mahmood, Christopher Leckie, and Parampalli Udaya " An Efficient Clustering Scheme to Exploit Hierarchical Data in Network Traffic Analysis ". In *IEEE Transactions on Knowledge and Data Engineering*, Vol 20, No 6, June 2008.
- [2] Cao J., D. Davis, S. Vander Weil, and B. yu, "Time- Varying Network Tomography", *J. Am. Statistical Assoc.*, 2000.
- [3] Chadi Barakat, Patrick Thiran, Gianluca Iannaccone, Christophe Diot, and Philippe Owezarski. Modeling internet backbone traffic at the flow level. *IEEE Transactions on Signal Processing*, 51(8):2111–2124, August 2003.[2] Graham Cormode, S.Muthukrishnan, and Irina
- [4] Dai , B.-R. J.-W. Huang, M.-Y. Yeh, and M.-S. Chen, "Adaptive Clustering for Multiple Evolving Streams," *IEEE Trans. Knowledge and Data Eng.*, vol. 18, no. 9, pp. 1166–1180, Sept. 2006.
- [5] Goldschmidt O., "ISP Backbone Traffic Inference Methods to Support Traffic Engineering", *Proc. Internet Statistics and Metrics Analysis Workshop (ISMA '00)*, Dec. 2000.
- [6] Guha S, A. Meyerson, N. Mishra, R. Motwani, and L. O'Callaghan, "Clustering datastreams: Theory and practice," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 3, pp. 515–528, 2003.

- [7] Guha S., N. Mishra, R. Motwani, and L. O'Callaghan, "Clustering Data Streams", 41st Annual Symposium on Foundations of Computer Science, 2000.
- [8] Henzinger M. R. , P. Raghavan, and S. Rajagopalan, "Computing on Data Streams", External Memory Algorithms, Boston: American Mathematical Society, 1999.
- [9] Huang Z, M. Ng, H. Rong, and Z. Li, "Automated Variable Weighting in k-Means Type Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp.657-668, May 2005.
- [10] Hun-Jeong Kang, Hong-Taek Ju, Myung-Sup Kim and James W. Hong, "Towards Streaming Media Traffic Monitoring and Analysis", APNOMS 2002, September 2002, Jeju, Korea.
- [11] Jacobus van der Merwe, Ramon Caceres, Yang-hua Chu, and Cormac Sreenan, "mmdump- A Tool for Monitoring Internet Multimedia Traffic," ACM Computer Communication Review, Vol. 30, No. 5, 2000.
- [12] Jensen C , D. Lin, and B.C. Ooi, "Continuous Clustering of Moving Objects," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 9, pp. 1161-1173, Sept. 2007.
- [13] Kuman A, M. Sung, J. Xu, and J. Wang, "Data Streaming Algorithms for Efficient and Accurate Estimation of Flow Size Distribution", *Proc. ACM SIGMETRICS*, 2004.
- [14] Lakhina A, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk, and N. Taft, "Structural Analysis of Network Traffic Flows", *Proc. ACM SIGMETRICS '04, June 2004*.
- [15] Lakhina A, M. Crovella, and C. Diot, "Characterization of Network- Wide Anomalies in Traffic Flows", Technical Report BUCS-2004-020, Boston Univ., 2004.
- [16] Lockwood J. W, S. G. Eick, J. Mauger, J. Byrnes, R. P. Loui, A. Levine, D. J. Weishar, and A. Ratner, "Hardware Accelerated Algorithms for Semantic Processing of Document Streams", 2006 IEEE Aerospace Conference, March 4-11, 2006.
- [17] Lockwood J. W, S. G. Eick, D. J. Weishar, R. P. Loui, J. Moscola, C. Kastner, A. Levine, and M. Attig, "Transformation Algorithms for Data Streams, 2005 IEEE Aerospace Conference", March 5-12, 2005.
- [18] Li W , W.K. Ng, Y. Liu, and K.-L. Ong, "Enhancing the Effectiveness of Clustering with Spectra Analysis," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 7, pp. 887-902, July 2007.
- [19] Madina A , K. Salamatian, N. Taft, I. Matta, and C. Diot, "A Two Steps Statistical Approach for Inferring Network Traffic Demands", revision of Technical Reports BUCS- TR-2003-003, Mar. 2004.
- [20] Maulik U and S. Bandyopadhyay, "Performance Evaluation of Some Clustering Algorithms and Validity Indices," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1650-1654, Dec. 2002.
- [21] Mouratidis K , D. Papadias, S. Bakiras, and Y. Tao, "A Threshold-Based Algorithm for Continuous Monitoring of K Nearest Neighbors," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 11, pp. 1451-1464, Nov. 2005.
- [22] Patrikainen A and M. Meila, "Comparing Subspace Clusterings," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 7, pp. 902-916, July 2006.
- [23] Su M.-C. and C.-H. Chou, "A Modified Version of the k-Means Algorithm with a Distance Based on Cluster Symmetry," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 674-680, June 2001.
- [24] Subhabrata Sen and Jia Wang, "Analyzing Peer-to-Peer Traffic Across Large Networks", IMW2002 Workshop, 2002, Marseille, France.
- [25] Vijayakumar M and R.M.S.Parvathi. " Concept Mining of High-Volume Data Streams in Network Traffic using Hierarchical Clustering". In *Proceedings of the Fourth International Conference, Amrutvani College of Engineering, Sangamner, Maharashtra. Proc. ITECH '09* pp 69, March 2009.
- [26] Wang J, D. Miller, and G. Kesidis, "Efficient Mining of the Multidimensional Traffic cluster Hierarchy for Digesting Visualization, and Anomaly Identification", *IEEE J. Selected Areas of Comm.* Vol. 24, no. 10, pp. 1929-1941, Oct. 2006.
- [27] Xu R and D. Wunsch II, "Survey of Clustering Algorithms," IEEE Trans. Neural Networks, vol. 16, no. 3, pp. 645-678, 2005.
- [28] Yip K.Y.L., D.W. Cheng, and M.K. Ng, "HARP: A Practical Projected Clustering Algorithm," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 11, pp. 1387-1397, Nov. 2004.

AUTHORS PROFILE



M. Vijayakumar has completed his Bachelor of Engineering in Computer Science in Bharathiar University, Tamilnadu, India and Master of Engineering , in Computer Science in Anna University, Chennai , Tamilnadu, India. He has started his teaching profession in the year 2004 in Muthayammal Engineering College, Tamilnadu. India., At present, he is working as an Assistant Professor in the department of Computer Science and Engineering in Sasurie College of Engineering, Tamilnadu. India. He has published 15 research papers in National and International journals and conferences. Currently he is a part time Research Scholar in Anna university of Technology, Coimbatore. His areas of interest are Data mining, Knowledge Engineering, Clustering algorithms and Network security. He is a life member of ISTE.



Dr. R.M.S. Parvathi has completed her Ph.D., degree in Computer Science and Engineering in 2005 in Bharathiar University, Tamilnadu, India. Currently she is a Principal and Professor , Department of Computer Science and Engineering in Sengunthar College of Engineering for Women, Tamilnadu, India, She has completed 20 years of teaching service. She has published more than 28 articles in International / National Journals. She has authorized 3 books with reputed publishers. She is guiding 20 Research scholars. Her research areas of interest are Software Engineering, Data Mining, Knowledge Engineering, and Object Oriented System Design. She is a life member of ISTE and Indian Computer Society.

A.R.Q. techniques using Sub-block retransmission for wireless networks

A.N.Kemkar, Member, ISTE and Dr. T.R.Sontakke Member,ISTE

Abstract—In this paper we mainly focus our investigation on the throughput performance in conjugation with sub-block transmission scheme. The throughput of a wireless data communications system depends on a number of variables, one of it is length of the message blocks. Over a noisy communication medium like wireless medium used for mobile ad-hoc network, our propose scheme performs effectively. In propose scheme random length of the message is divided in to fixed length blocks and applying ARQ techniques if the error occurs. A threshold model is used for fading channel, estimation and CRC detection codes are used. Comparison of transmission efficiency of proposed scheme with varying channel condition is shown.

Index Terms— FEC, Hybrid ARQ, BER.

1. INTRODUCTION:

Wireless channels are highly affected by unpredictable factors such as co-channel interference, adjacent channel interference, propagation path loss, shadowing and multi path fading. The unreliability of media degrades the transmission quality seriously. Automatic Repeat ReQuest (ARQ) and Forward Error Correction (FEC) schemes are frequently used in wireless environments to reduce the high bit error rate of the channel.

As we have seen, the throughput efficiencies of all the basic ARQ schemes are functions of the packet size n . [1],[2],[3],[4],[5]. Our main result is a mathematical technique for determining the block size as a function of the other variables like BER, signal-to-noise ratio.

A.N.Kemkar¹, S.R.T.M.U, Nanded.
+91-9819150392, ankemkar@gmail.com

Dr.T.R.Sontakke²
Ex.Director – S.G.G.S.I.T.E.- Nanded
Principal, Sidhant college of Engineering
Pune.+91-
9822392766, trsontakke@gmail.com

In an attempt to improve throughput performance, we have included an analysis using forward error correcting (FEC) block codes (used in Hybrid ARQ). The optimum amount of FEC coding was found to be dependent upon the Block length. As the Block length increases, the number of correctable errors to optimize the throughput also increases, mathematical expression is shown in 2.2.

The paper is organized as follows. In Section 2 .Summary on the Related work and basic concept. In section 3.scheme description and system model. We consider the performance analysis of the proposed scheme for simulation study in section 4. followed by conclusion in Section 5

2.RELATED WORK :

2.1: Related Work- The efficiency of HARQ scheme is compared with GBN schemes using different lengths of IP Blocks. Further show that usage of smaller Blocks and hybrid schemes leads to an improved throughput. Differences between pure and hybrid GBN schemes are also discussed. [1]

When the channel is quiet the sub-block retransmission scheme behaves like a conventional ARQ or hybrid ARQ scheme. As the channel becomes increasingly noisy, the data block is divided into smaller sub-blocks for transmission. Each sub-block is encoded for error control by an appropriate shortened code of which the code length is adapted to the corresponding channel BER.[2] Further optimum block size in accordance with the channel conditions [4] A single code HARQ scheme was proposed in which transmitter is operating in any one mode with the degree of errors encounter. The operating state is selected based on the channel BER. Data bits are divided in blocks and are encoded with shortened codes. During the retransmission new coded blocks are combined and at the receiver end proper decoding techniques are used to separate retransmitted blocks from the new blocks.

2.1. BASIC CONCEPT -

Analytical expression how Throughput performance of the system varies with the size of the block length and FEC. Consider the following two cases to verify the Throughput performance of the system.

Case 1. Blocks are transmitted without FEC.

Case 2. Blocks are transmitted with FEC.

Our analysis includes the following simplifying assumptions: 1. The CRC decoder detects all errors in the output of the FEC decoder. 2. Transmission of acknowledgments from the receiver to the transmitter is error free and instantaneous.

System throughput (T) is the number of payload bits per second received correctly:

$$T = \frac{K}{L} R f(\gamma) \quad (1)$$

where $\left(\frac{KR}{L}\right)$ b/s is the payload transmission rate

Where $f(\gamma)$ = Block success rate defined as probability of receiving block correctly. Probability is a function of signal to noise ratio.

$$\gamma = \frac{E_b}{N_o} \quad (2)$$

In which $E_b = \frac{P}{R}$ joules in received energy per bit.

where R = Transmission rate. Probability is a function of signal to noise ratio.

$$\gamma = \frac{E_b}{N_o} = \frac{P}{N_o R} \quad (3)$$

Each Block, of length L bits, is a combination of a payload (K) and overhead $(L-K)$. Because the Block success rate, $f(\gamma)$ is a decreasing function of, L there is an optimum Block length, L^* . When

$L \ll L^*$ excessive overhead in each Block limits the throughput. When $L > L^*$ Block errors limits the throughput.

For case 1. When there is no forward error correction coding. In this case

$$f(\gamma) = (1 - P_e(\gamma))^L \quad (4)$$

where $f(\gamma)$ = block success rate defined as probability of receiving block correctly.

$P_e(\gamma)$ is block error rate.

Therefore, in a system without FEC, the throughput as a function of L , from (1)

$$T = \frac{L-C}{L} R (1 - P_e(\gamma))^L \quad (5)$$

Case 2.: Now instead of transmitting those L bits with no error correction capability, we will now add B error correcting bits and transmit a total of bits $L+B$. Using a block code forward error correction scheme, the minimum number of B bits required to correct t errors is given by [5]

$$B \geq \log_2 \left[\sum_{n=0}^t \binom{L+B}{n} \right] \quad (6)$$

Now that we can correct ' t ' errors, our block success rate, $f(\gamma)$ should be larger than its previous value with no error correction. Recall that $f(\gamma)$ with $t=0$ is given by:

$f(\gamma) = (1 - P_e(\gamma))^L$ where $P_e(\gamma)$ is the probability of a bit error as a function of the SNR. Now, with error correction capability, the Block success rate for some arbitrary value of t is [7]

$$f_t(\gamma) = \sum_{n=0}^t \binom{L+B}{n} P_e^n(\gamma) (1 - P_e(\gamma))^{L+B-n} \quad (7)$$

Our new equation for the throughput as a function of the signal to noise ratio is:

$$T(\gamma) = \left(\frac{L-C}{L+B} \right) \left(\frac{\frac{P}{N_o}}{\gamma} \right) f_t(\gamma) \quad (8)$$

From (5) and (8) it is clear that throughput of the system is a function of message block length. Further (5) and (8) are used for pure and hybrid ARQ techniques.

3. SCHEME DISCRIPTION AND SYSTEM MODEL :

This paper presents a sub-block retransmission scheme for ARQ. The data block is divided into smaller sub-blocks for transmission. Each sub-block is encoded by an appropriate error detection codes. The encoded block is then transmitted. The received block is checked for errors sub-block by sub-block. The proposed scheme provides improved throughput by retransmitting only the sub-blocks in the occurrence of errors.

3.1 SYSTEM MODEL :

We consider an ad-hoc network with V nodes and assume that each node is equipped with only one antenna. A Point to Point protocol is used at the medium access control layer. A Selective repeat request ARQ mechanism is used. Particularly, the source node transmits a data packet with a C -bit CRC attached. The destination node detects CRC and then sends an acknowledgement that is either positive (ACK) or negative (NACK) back to the source node. If the packet is correctly detected by the destination node (with ACK feedback), the source node continues to transmit a new data packet and the above process is repeated. Otherwise, retransmission will start. A threshold model for channel characterization is used for fading channel.

4. PERFORMANCE ANALYSIS OF THE PROPOSED SCHEME:

The performance analysis of the scheme is measured in terms of throughput of the proposed scheme. Further we show the comparison of throughput with sub block and without sub block transmission schemes.

Expression of throughput for ARQ for present scheme:

$$\eta = \frac{K}{E[T]} \quad (9)$$

where K =information bits in a block. $E[\cdot]$ =Expectation of number of transmitted bits in a given block.

$$T = Mn + \sum_{i=1}^{\infty} T_i \quad (10)$$

where M =number of sub blocks, n =number of bits in a sub block, T_i =number of transmitted bits for i^{th} transmission.

$$E[T] = Mn + \sum_{i=1}^{\infty} E[T_i] \quad (11)$$

where $E[T]$ =Average number of transmitted bits.

Out of M sub blocks if L sub blocks are transmitted at the i^{th} retransmission, then random variable, T_i takes the value Ln , if L out of M sub-blocks are retransmitted at the i^{th} retransmission.

SIMULATION RESULTS: We evaluate the performance of the proposed scheme implemented with Matlab. We run the simulation for two schemes i.e. with sub block transmission and without sub block transmission. The simulation parameters are shown in the table 1. Simulation run for 5000 total blocks. Result is the average of independent experiments where each experiment uses different randomly generated uniform parameters. We use mean values which are obtained independent experiments as a basic data to get the result.

Simulation results are shown Table 2

Table 1: System Parameters:

Parameters	Notation	Values
Signal to Noise Ratio	γ	Varied
Total number of blocks	-----	5000
Total sub block	M	32
Information bits in a block	K	16
Packet length	n	5000*32*16
Max. number of Retransmissions	-----	3
Number of sub blocks retransmitted	L	Varied
Cyclic Redundancy Check	CRC	Varied
Bit error rate	BER	Varied
Packet error rate	PER	Varied
Throughput efficiency	η	Varied

Table 2 : Simulation Results: Following Simulation results shows the comparison of Throughput efficiency verses varied block size verses changing channel condition in terms of PER.

Table 2:

Block length		Packet error rate		Throughput	
Without sub block	With sub block	Without sub block	With sub block	Without sub block	With sub block
Whole block is transmitted with out sub division.	Whole block is divided into sub 4 blocks	0.1	0.1	0.9	0.99
		0.3	0.3	0.86	0.96
		0.5	0.5	0.66	0.94
		0.7	0.7	0.59	0.93
		0.9	0.9	0.57	0.9
		1	1	0.5	0.89

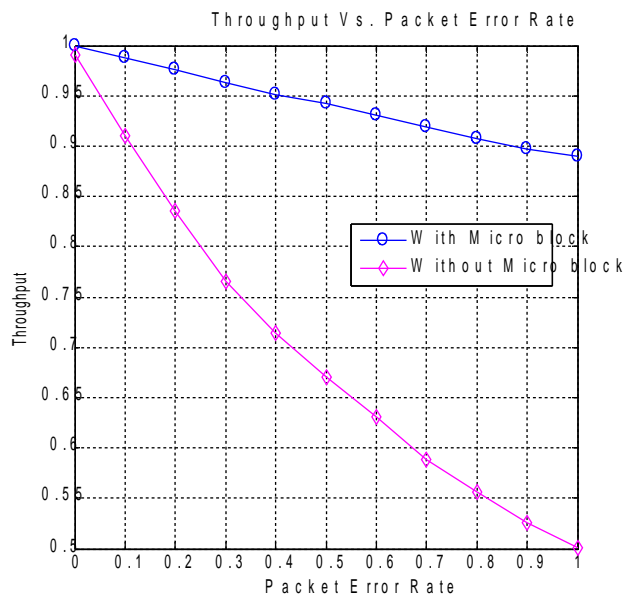
5. CONCLUSION:

From the Fig.1 shown below performance of the proposed scheme. This paper has presented a sub-block retransmission scheme for

1. With micro block ARQ
2. Without micro block ARQ

Proposed sub-block retransmission schemes showed better overall performance compared to the other competitive schemes by retransmitting without sub-blocks in the occurrence of errors. From Table 2 and the Fig.1 it is clear that as the packet error rate increases i.e. channel condition gets deteriorated the throughput performance goes down by more than 40% as compare to proposed scheme. There fore we proposed sub-block retransmission scheme with ARQ is more reliable than existing i.e. without sub block transmission scheme.

Fig.1: Performance of proposed scheme vs. existing schemes.



6. REFERENCES:

- [1] W.W. Chu. Optimal message block size for computer communications with error detection and retransmission strategies. *IEEE Transactions on Communications*, COM-22:1516– 1525, October 1974.
- [2] J.S. DaSilva, H.M. Hafez, and S.A. Mahmoud. Optimal packet length for fading land mobile data channels. In *Proceedings of ICC 1980*, pages 61.3.1– 61.3.5, June 1980.
- [3] R.L. Kirlin. Variable block length and transmission efficiency. *IEEE Transactions on Communication Technology*, COM-17:350–355, June 1969.
- [4] E. Modiano. An adaptive algorithm for optimizing the packet size used in wireless ARQ protocols. *Wireless Networks*, 5:279–286, July 1999.
- [5] J.M.Morris. Optimal block lengths for ARQ error control schemes. *IEEE Transactions on Communications*, COM-27:488–493, February 1979

Performance Analysis of Delay in Optical Packet Switching Using Various Traffic Patterns

A.Kavitha/Chettinad College of Engineering & Technology

IT dept
Chettinad College of Engineering & Technology,
Karur, Tamilnadu, India

V.Rajamni/Indra Ganesan College of Engineering,

Indra Ganesan College of Engineering
Trichy, Tamilnadu, India

P.Anandhakumar/Madras Institute of Technology,

IT Dept
Chennai, Tamilnadu, India

Abstract— Quality of Service parameters are improved for development of optical packet switching technology. Delay is an important parameter in optical packet switching networks and it affects the performance of the network. In this paper, a mathematical model is presented to evaluate the delay rate. Delay rates are analyzed for fixed packet length and variable length packet for various traffic patterns viz. Non-uniform, Poisson and ON-OFF traffic models for various service classes using Reservation Bit technique. The results are compared with the existing port based First-Fit wavelength assignment algorithm. Here delay rates are reduced by 29% in our class based model than the port based model.

Keywords—component; Optical Packet Switching (OPS), RB (Reservation Bit algorithm), FF (First-Fit Wavelength assignment algorithm), Quality of Service (QoS), Packet Loss Rate (PLR), BER (Bit Error Rate), WDM (Wavelength Division Multiplexing).

I Introduction

Due to the explosive growth of internet applications in recent years, data traffic has been exceeded the telephony traffic and bandwidth demands have been continuously increasing. It is also expected of the future networks to transport heterogeneous traffic services including multimedia and interactive applications necessitating bandwidth guarantees, minimum delay, less PLR, controlled jitter and etc. QoS provisioning seems therefore a mandatory task. Optical networks offer an extremely high traffic bandwidth capable of providing communication channels for several hundred nodes. Thus, the network traffic requires the network to evolve by increasing transmission capacity of optical fibers as well as switching capability. There are three switching schemes in optical networks namely optical circuit switching, optical burst switching (OBS) and optical packet switching (OPS). In the optical circuit switching, a dedicated end-to-end light path is established for each connection. Thus the transmission delay can be guaranteed

and there is no virtually any loss, but less utilization of wavelength in this technique. In OBS, data is sent in bursts and a burst control message is sent ahead of each data burst to reserve a wavelength at each hop based on the expected arrival time of the data burst. In OPS, messages are transmitted in packets. At each switching node, the packet head is processed in the electrical domain for routing purpose and the packet data is kept in the optical domain. Wavelengths can be efficiently used in OPS [1]. Thus the optical packet switching has emerged as one of the most promising technologies for future telecommunication networks. OPS utilize very high bandwidth in the optical fiber using WDM. WDM offers an aggregate throughput of the order of terabits per second. WDM is widely becoming accepted as a technology for meeting growing bandwidth demands, and WDM systems are beginning to be deployed in both terrestrial and undersea communication links. Thus, WDM offers an excellent platform for carrying IP traffic. Consequently, OPS technology has many advantages, attract more intensive attention than ever. The next generation telecom infrastructure definitely comprise of optical networks with improved QoS.

In order to improve the performance of QoS in optical packet switching network, a detailed study has been made in this paper. The performance analysis of optical packet switching consists of two important issues namely packet loss and delay. In order to provide better QoS in optical packet switching, PLR and delay should be reduced. Packet loss and delay are not new issues in optical networks; however minimum loss and delay provide better QoS in Optical packet switching. In our earlier report the performance of 8B/10B code, Systematic code and Viterbi code in optical transmission in terms of Bit Error Rate has been analyzed. In the physical layer, transmission is done on bit by bit basis and Bit Error Rate has been reduced in the physical layer. When the bit errors in the physical layer are rectified, it will reduce the packet loss rates in the higher layers [2,3]. We have already reported that PLR has been

reduced in non-uniform traffic pattern [4]. We have also implemented a RB algorithm for minimizing PLR in buffered non-uniform traffic pattern of optical packet switching mechanism [5].

In this paper we have analyzed the asynchronous OPS in terms of delay. Delay rate for fixed length packet and variable packet length has been studied for various traffic patterns viz. Non-uniform, Poisson and ON-OFF traffic models.

This paper has been organized as follows: Section II describes the architecture of asynchronous OPS. In section III, the analysis of delay rate is carried out for various traffic patterns in asynchronous OPS. In section IV, results along with discussions are presented and Section V deals with the conclusion and future work.

II. Description of Architecture

The architecture used in this paper is presented in [4,5]. It has been reproduced for the reference. The size of switch under consideration is $N \times N$. The switch has F input fibers and F output fibers. By utilizing WDM, each fiber provides N wavelengths to transport data with a capacity of C bps. Buffers with the size of 5 are used in the OPS switch catering to each of the service classes. The switching process in OPS can take one of the two main forms. It can be synchronous (time slotted) with the fixed packet length or asynchronous (non-slotted) with variable packet lengths. In synchronous operation mode, all arriving packets have a fixed size and their arrival on each wavelength is synchronized on a time-slot basis, where a time slot is the time needed to transmit a single packet.

The operations of optical packet switching can be briefly described as follows: When a packet arrives at the switch, the packet header is extracted and processed electronically by the control module. While the header is processed, the packet payload is buffered in the optical domain using FDL processing buffers. Based on the destination, information extracted from the packet header and the control module decides to which output fiber/wavelength the packet is switched and configures the switch accordingly. Contention occurs when two or more packets are assigned to the same output port on the same wavelength at the same time [6 and 7]. The network has d service classes, ranging from service class 0 to service class $d-1$. We assumed that the output on a single fiber/wavelength as the tagged fiber/wavelength. Delay is calculated for class i traffic at the tagged output fiber. Fig1 shows a switch for packet arrivals to a tagged output fiber must originate from one of the FN input wavelengths.

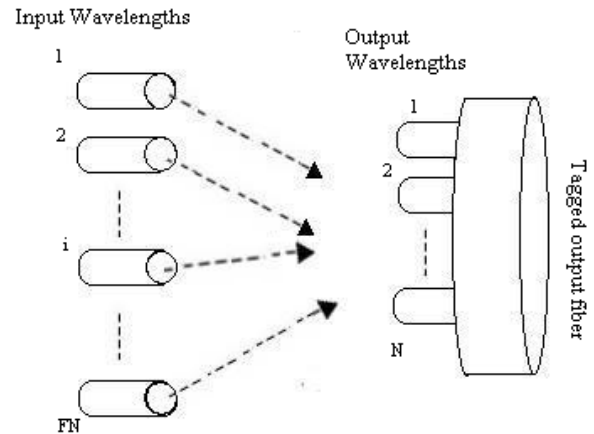


Fig 1. Model of the switch in OPS network under consideration

The following assumptions are made for the simulation:

Let j_i denote the number of class i packets that arrive in a time slot. The total number of packet arrivals at the fiber in a time slot is k . That is $j_0 + j_1 + \dots + j_{d-1} = k$. In order to isolate the service classes, the parameter l_i ($0 \leq l_i \leq N$) is introduced, which is the number of wavelengths reserved for Class i traffic in case of contention in a time slot. For a service class i , if $j_i < l_i$ (the number of incoming packets are less than the wavelengths assigned), it will result in $j_i - l_i$ free slots. For a service class i , if $j_i > l_i$ (the number of incoming packets are greater than the wavelengths assigned), resulting in $l_i - j_i$ overflow of packets.

The RB algorithm [5] is implemented by introducing buffers. The RB algorithm is used for slotted OPS wherein buffers are used to avoid overlapping of packets. This algorithm is designed and used to improve the QoS of the networks in terms of reduction in PLR. By using the same algorithm delay rate is found and analyzed in slotted OPS with fixed packet lengths for various traffic patterns. Hence QoS is improved in terms of reduction in PLR and delay [4,5]. In this paper the same algorithm is implemented in asynchronous OPS for the packets of variable lengths.

III. Operating Principle

Operation of optical packet switch in a synchronous manner with fixed length of packets is explained in [5]. Here we assume that optical packet switch operates in an asynchronous manner with variable length of packets. Analysis of delay rate for variable packet sizes in asynchronous OPS using three types of traffic patterns viz. Non-uniform, Poisson and ON-OFF traffic models is presented in this paper. In contrast to [8], immaterial of packet size, allotted time slots remain the same and no shifting of time slot allocation is done.

In non-uniform traffic pattern all nodes are not to receive and send similar volumes of traffic [4, 9]. The number of packets arrived and transmitted of packets is not equal. There is an incoming packet for every slot. We assume that

each packet has equal probability of $1/N$ being addressed to any given output port and successive packets that arrive at the tagged output/wavelength are independent. Also, the packet arrival time at the input and transmission time to the output port are not equal. This is due to randomness of packet arrival. This increases the occupancy of the free slots. In Poisson traffic model, packet arrivals are random and are mutually independent. The Poisson distribution can be obtained for packet arrivals during an infinitesimal short period of time λt , where λ is called the arrival rate. Packet arrival at the output fiber is a single Poisson process with λ . In Poisson model, the number of arrivals in non-overlapping intervals is statistically independent. The arrival rate λ is expressed as the average number of arrivals during a unit of time. The time distance between consecutive packet arrivals is exponentially distributed [10]. However Poisson arrival model is not an accurate model considering packet arrivals at the tagged output fiber/wavelength in OPS. The Poisson arrival model assumes an infinite number of sources, which is not the case in a real switch. When using a Poisson arrival model, there is the possibility of independency between the service time and the packet inter-arrival time [11].

Bursty traffic is also known as ON-OFF traffic model. Both the ON and OFF periods are distributed using exponential distribution with the mean $1/\mu$ and $1/\lambda$ respectively. It is an alternating process where an OFF period follows an ON period, and an ON period follows an OFF period. During ON periods, series of packets are transmitted from the source node and the time is called active periods. OFF periods are called passive periods and no packets are transmitted from the source. Active periods of a source are exponentially distributed with one specific mean value, and passive periods are exponentially distributed with another mean value. During an active period, packets are generated at regular periods. The most commonly used VOIP traffic model is based on a two-state on-off model of a single voice source. When a voice source is transmitted, it is in the ON state, and when the voice source is silent, it is in the OFF state; and the ON and OFF states appear alternatively. ON-OFF sources, each of which exhibits a phenomenon called the "Noah Effect" resulting in self similar aggregate traffic. The Noah Effect for an individual ON-OFF source model results in ON and OFF periods, i.e. "train lengths" and "intertrain distances" that can be very large without negligible probability [12 and 13]. The number of packet arrivals divides the timeslot and produces the time slice which is utilized for packet transmission. Hence the packet arrival rate is random, time slice is determined according to randomness of the packet arrival. ON and OFF periods are used with their own mean values and using the same mean values, various packets arrive and are transmitted during ON period and idling occurs during the OFF period. During ON period, packet is transmitted and there is no flow of packet during OFF period. In this model, FN independent state model generate packet arrivals at a tagged output fiber. In

this paper, at every node there should be arrival(s) of different classes of packets. In ON-OFF model, ON periods and OFF periods occur at each node for different classes. When one class is in ON period, other classes may be in OFF period at that particular node and OFF period of one class is used by the other classes for transmitting packets. Also, other possibilities are (1) all classes may be in OFF period, (2) all classes may be in ON period and (3) some of the classes may be in ON period and rest may be in OFF period.

We assume that the switch is having buffers. In RB algorithm, the packets use the wavelength according to their service classes. If there is flow of packets into the ports with a specified service class, the incoming packets with assigned wavelengths occupy the ports as per the assigned service class. When the assigned wavelengths in one service class are occupied, it checks for the free wavelengths of other service classes and the packets will occupy the free slots in the other service classes. When the assigned wavelengths are completely occupied in all the ports, the packets overflow. By introducing buffers, the packets that overflow are saved. When a free slot is not available for an incoming packet, instead of dropping that packet, it will be saved in the buffer which is provided in the switch. Before entering the buffer, a bit is added in the packet header for the purpose of reservation with respect to their service classes. Whenever free slots are available, the packets in the buffer occupy free slots. Buffered packets will have the priority over the incoming packets. In the FF algorithm, all wavelengths are numbered in a certain order, for example ascending order from 0 to $W-1$, where W is a number of wavelengths. When the deciding port attempts to assign a wavelength, it sequentially searches all wavelengths in an ascending order and assigns the first available wavelength [14]. In class based model, each node transmits the packets according to their classes. Buffers are placed in the port for every class. In port based model, wavelengths are placed in a sequential order. Irrespective of the class, the available wavelength is used by the incoming packets in a sequential order and buffers are only placed in each port.

FF algorithm is implemented in port based and packets are transmitted according to their wavelengths, whereas transmission of packets is class based in RB algorithm and packets are transmitted according to their service classes and wavelengths. Thus the drop rates of packets are reduced in optical networks resulting in improved QoS. The delay rate is found and analyzed in asynchronous OPS for various traffic patterns viz. Non-uniform, Poisson and ON-OFF traffic models for various service classes with packets of variable lengths. The delay rates for RB algorithm and FF algorithm are found and compared.

For fixed size packets and variable size packets in OPS, delay rate is encountered by implementing RB algorithm. The packet includes payload and header. Fig 2 shows the packet header. For fixed size packets, size considered is 512 bytes. For variable packet size, the range of packet size is in the

range of 512 bytes and 2k bytes. 20 bytes is included as header along with the payload invariably for any type of packets. Variable size packets are used in VoIP applications. In variable packet size, the size has been controlled by the application. Packet sizes of the range of 1024 to 2048 bytes show good efficiency in terms of bandwidth and reliability in Digital Video Broadcasting [15]. Packet size is measured using uniform min and max distribution. Packet size is chosen depending on the application.

Source IP address	Destination IP address	Source Port	Destination Port	Packet Sequence no	Time Stamp	Flow id
4 bytes	4 bytes	2 bytes	2 bytes	2 bytes	4 bytes	2 bytes

Fig. 2 Structure of the packet header

When a packet is send from one node to other, the following delays occur: (1) transmission delay (time required to send all bits of packet into the wire), (2) propagation delay (the time taken by the packet to travel through the wire), (3) processing delay (the time taken to handle the packet in the network system), and (4) queuing delay (the time taken is buffering the packet before it can be sent). In most cases, the delay (2) and (4) are considered in simulations and measurements. The transmission delay (1) is usually small for fast links and small packets and is therefore not considered. Traditionally, the processing delay (3) has also been negligible [16]. In our measurements, we consider the delay (2) and (4). The mean delay for the above said traffic pattern is found using equation (1).

$$T_D = T_{\text{Propagation}} + T_{\text{Queue}} \quad (1)$$

Initially, we calculate the propagation delay that occurs when a packet travels from the source to the destination. Next, the queue delay experienced by a packet is calculated using equation (2). This delay is due to waiting period of the packet in the queue. A packet is in a queue, if a free wavelength is not available at that particular time slot.

$$T_{\text{Queue}} = \frac{1}{N} \sum_{i=0} (T_i) \quad (2)$$

where T_i is the transmission time of class i packets at particular time slice.

Summation of the waiting time in the buffer and the transmission time between source node and destination node through the switch is considered as delay and the same is found for the above said traffic patterns.

IV. Results and Discussions

The delay values for the fixed length packets in slotted OPS using RB algorithm is studied and is also compared to the FF algorithm in our earlier paper [5]. In this paper, a detailed analysis is carried out to find delay in asynchronous OPS for variable length packet and is compared to FF algorithm.

We consider 240 packets for the simulation purpose. Delay for class i packets are calculated in the tagged fiber. Wavelength assigned is 16 and total time slot chosen is 10, hence this architecture can transmit 160 packets in a time

slot. Buffers are used along with RB technique, 240 packets are chosen with an Erlang load of 1.5.

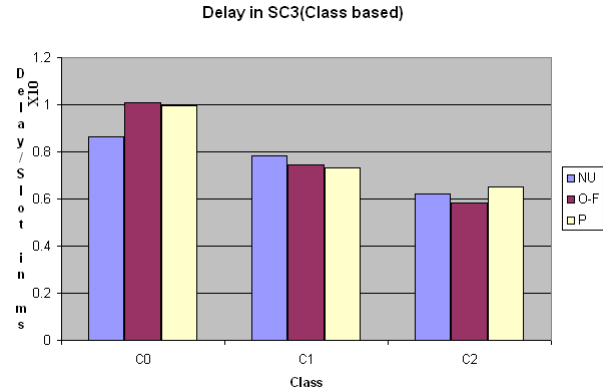


Fig 3. Delay rate for various traffic patterns for service class 3 using reservation bit technique.

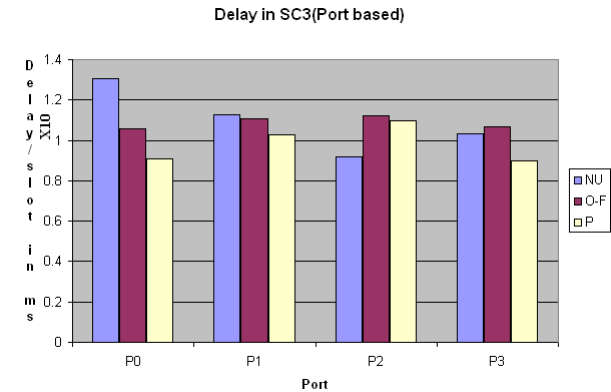


Fig 4. Delay rate for various traffic patterns for service class 3 using First-Fit wavelength assignment algorithm

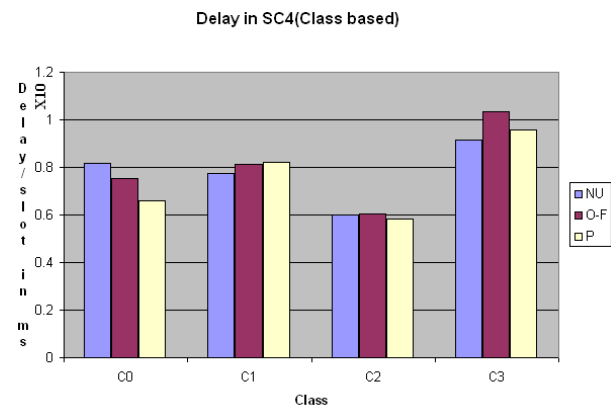


Fig 5. Delay rate for various traffic patterns for service class 4 using reservation bit technique

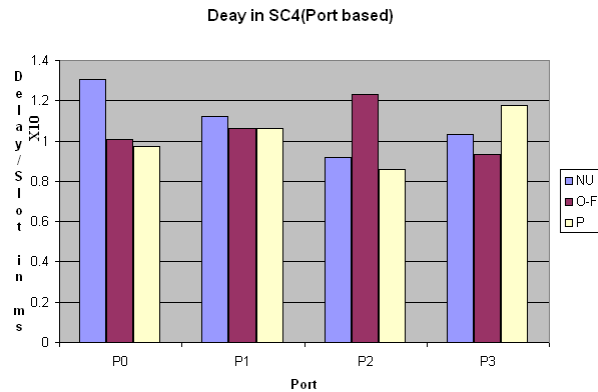


Fig 6. Delay rate for various traffic patterns for service class 4 using First-Fit wavelength assignment algorithm

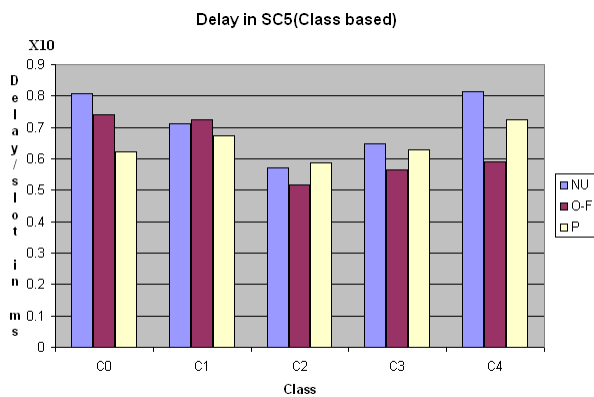


Fig 7. Delay rate for various traffic patterns for service class 5 using reservation bit technique

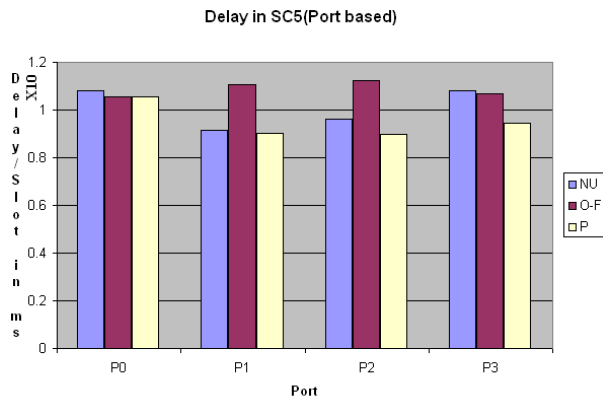


Fig 8. Delay rate for various traffic patterns for service class 5 using First-Fit wavelength assignment algorithm

Delay Rate for various Traffic pattern

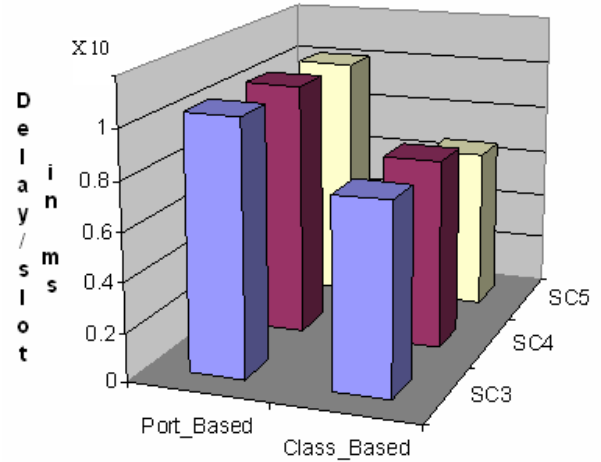


Fig 9. Delay rate comparison Transmitted Bytes

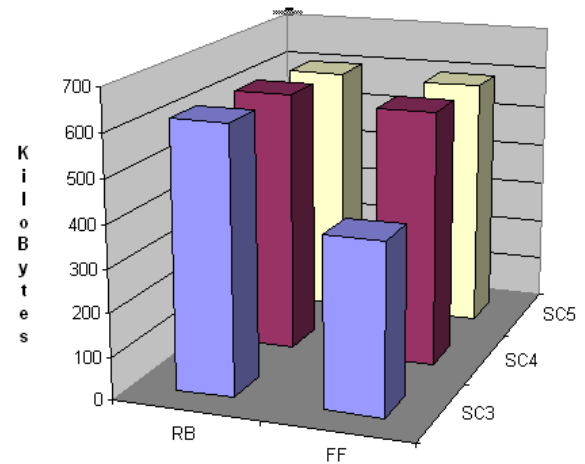


Fig 10. Comparison of data transmission

In service class 3, RB technique has 15 buffers, each service class has 5 buffers, but in FF algorithm, each port has 5 buffers and the total number of buffers is 20. 10.96ms, 10.88ms and 9.83ms are the delay values while employing FF algorithm and 7.55ms, 7.79ms and 7.91ms are the delay values while employing RB technique in asynchronous OPS using Non-Uniform, ON-OFF and Poisson traffic model respectively for Service class 3 and the same is shown in figs 3 and 4.

In Service class 4, both the techniques have 20 buffers. 10.94ms, 10.56ms and 10.18ms are the delay values while employing FF algorithm and 7.77ms, 8.01ms and 7.561ms are the delay values while employing RB technique in asynchronous OPS using Non-Uniform, ON-OFF and Poisson traffic model respectively for Service class 4 and

the same is shown in figs 5 and 6. From these it is seen that both the techniques have same amount of buffer. FF algorithm exhibits more delay, but delay is less in our approach.

While in service class 5, RB techniques have 25 buffers and FF algorithm need 20 buffers. 10.09ms, 10.88ms and 9.50ms are the delay values while employing FF algorithm and 7.1ms, 6.27ms and 6.48ms are the delay values while employing RB technique in asynchronous OPS using Non-Uniform, ON-OFF and Poisson traffic model respectively for Service class 5 and the same is shown in figs 7 and 8.

For all the service classes under consideration, buffers are more or less the same for RB technique and FF algorithm, whereas the delay rate is slightly higher in FF algorithm. The above statement is true when the class based transmission is compared with port based transmission.

Simulation results show that for all service classes under any type of traffic pattern, class based model produces 29% reduction of delay rate when compared to port based model and the same is shown in fig 9. In RB technique the buffered packets with respect to their classes will occupy the free slots of the corresponding service class wavelengths on First Come First Serve basis which reduces the waiting time in the buffer whereas in FF algorithm wavelength utilization is sequential, so buffered packets wait until they are serviced with the next sequential order of wavelengths. Thus delay rate is less in asynchronous OPS when RB technique is employed. The delays are improved much in our algorithm. This is due to the class based transmission.

Fig 10 shows the total number of transmitted bytes per slot. It is shown that more number of packets is transmitted in class based model for all service classes for all traffic patterns under consideration when compared to port based model. Hence class based model produces lesser delay as well as more number of bytes transmission when RB technique is employed. It is also shown that non-uniform traffic pattern produces lesser delay when compared to ON-OFF and Poisson model.

V. Conclusion

The delay rates for the Non-uniform, Poisson and ON-OFF traffic models for various service classes are analyzed. Minimum delay has been achieved in all service classes under consideration by using reservation bit technique. It provides lesser delay for all service classes for fixed and variable length packets when compared to first-fit wavelength assignment algorithm. It is seen from our simulation, Poisson arrival model which is assumed in the analysis approximates a more realistic model wherein all input wavelengths are modelled as independent on/off processes with exponential holding time. Also we have presented a comparative study of reservation bit technique and first-fit wavelength assignment algorithm for synchronous and asynchronous OPS. It is concluded that in Optical packet switching reservation bit technique reduces delay; hence QoS is improved and at the same time Non-

uniform traffic pattern result in better Quality of service compared to ON-OFF model and Poisson model. At the same time ON-OFF traffic pattern have better QoS when compared to Poisson process if OFF periods of one service classes are more efficiently utilized by other service classes.

- [1] Biao Chen and Jianping, "Hybrid Switching and P-Routing for Optical Burst Switching Networks", IEEE Journal on Selected areas in communications, vol 21, no 7, pp.1071-1080, Sep 2003.
- [2] A.Kavitha, V.Rajamani and P.AnandhaKumar, "Performance analysis of Coding Techniques to find BER in Optical Transmission", IEEE 1st International Conference on Advanced Computing- ICAC 2009, 13-15 Dec. 2009 Page(s):21 - 27
- [3] A.Kavitha, V.Rajamani and P.AnandhaKumar, "Evaluation of BER in Optical Packet Switching using Various Coding Schemes" IEEE Transaction on Optical Communication and Networking (under review).
- [4] A.Kavitha, V.Rajamani, "Performance Analysis of Slotted Optical Switching Scheme using Non-Uniform traffic" – Journal of Optical Communications, Vol. 29, July 2008, pp.107-111.
- [5] A.Kavitha, V.Rajamani and P.AnandhaKumar, "Performance Analysis of Slotted Optical Packet Switching Scheme in Non-Uniform Traffic Pattern Using Reservation Bit Technique"- Selected for publication- INFOCOMP, Journal of Computer Science.
- [6] H. Overby and N. Stol, "Evaluating and comparing two different service differentiation methods for OPS: the wavelength allocation algorithm and the preemptive drop policy," in Proc. 3rd Int. Conf. Networking, vol. 1, Feb. 2004, pp. 8-15.
- [7] S.Bjornstad, N.Stol and D.R.Hjelme, "A highly efficient optical packet switching node design supporting guaranteed service" in proc. Of SPIE, Vol 4910, 2002, pp.63-74.
- [8] BO Wen, Ramakrishna Shenai, and Krishna Sivalingam, "Routing Wavelength and Time-Slot-Assignment: Algorithms for Wavelength-Routed Optical WDM/TDM Networks", Journal of Lightwave Technology, vol 23, no 9, pp.2598-2609, Sep 2005.
- [9] Adisak Mekittikul, Nick McKeown "Scheduling VOQ Switches under Non-Uniform Traffic", CSL Technical Report, CSL-TR 97-747, Stanford University, 1997 Stanford University, Stanford.
- [10] Wuyi Yue, Yutaka takahashi,hideaki tagaki, "Advances in Queueing theory and network application", Springer,ISBN: 978-0-387-09702-2, e-ISBN: 978-0-387-09703-9.
- [11] Harald Overby and N.Stol "Quality of Service in synchronous bufferless optical packet switched networks" Kluwer Telecomm. Sys. Vol 27, 2004. pp. 151- 179.
- [12] Mihails Kulikos and Ernests Petersons "Remarks Regarding Queueing Model and Packet Loss Probability for the Traffic with self – Similar Characteristics", Networks", International Journal of Computer Science 3:2, Spring 2008, pp. 85-90.
- [13] Eric W.M. Wong and Moshe Zukerman, "Bandwidth and Buffer Tradeoffs in Optical Packet Switching", Journal of Lightwave Technology, vol 24, no 12, Dec 2006, pp 4790- 4798.
- [14] Xuehong Sun, Yuunhao Li, Ioannis Lambadaris and Yiqiang Q.Zhao, Performance Analysis of First – Fit Wavelength Assignment Algorithm in Optical Networks", IEEE 7th International Conference on Telecommunications – ConTEL2003, June 11-13, 2003, pp. 403-409.
- [15] Vadakital, V.K.M. Hannuksela, M.M. Razaee, M. Gabbouj, M., "Optimal IP Packet Size for Efficient Data Transmission in DVB-H", Proceedings of the 7th Nordic IEEE Signal Processing Symposium, pp. 82-85, 7-9 June 2006.
- [16] Ramaswamy Ramaswamy, Ning Weng and Tilman Wolf, "Characterizing Network Processing Delay", Proc. in Globecom 2004, IEEE Communication Society, pp.1629- 1634.

AUTHORS PROFILE



A.Kavitha received the B.E degree in Electronics and Communication Engineering from Madurai Kamaraj University, India in 1997. The Master of Engineering degree in Computer and Communication Engineering from Anna University, Chennai in 2004 and currently pursuing Ph.D degree in Optical Communication Networks in Anna University, Chennai. At present she is working as a Senior Lecturer in Chettinad College of Engineering and Technology, Karur. She has published more than 10 papers in referred National and International conferences/Journals. Her area of interest includes Networks, Computer Architecture, Digital Communication and etc.



V.Rajamani received the B.E degree in Electronics and Communication Engineering from National Engineering College, Anna University, India in 1990. The Master of Engineering in Applied Electronics from Government College of Technology, Bharathiyar University Coimbatore in the year 1995 and the Ph.D. degree in Electronics Engineering from Institute of Technology, Banaras Hindu University, Varanasi in 1999. He started his carrier as Lecturer in Mohamed Sathak Engineering College from 1991 onwards. He has held various positions in various Engineering Colleges. At present he is working as a Principal in Indra Ganesan College of Engineering, Tiruchirappalli. He has completed a project under AICTE - RPS scheme successfully. He has published more than 80 papers in referred National and International Journals/ conferences. His area of interest includes Device modeling, VLSI, Image processing, Optical Communication and system.

P.Anandhakumar received the B.E degree in Electronics and Communication Engineering from University of Madras, India in 1994. The Master of Engineering degree in Computer Science and Engineering from Bharathiyar University in 1997 and Ph.D degree in Computer Science and Engg. from Anna University, Chennai in the year 2006. At present he is working as a Assistant Professor in Madras Institute of Technology, Chennai. He published more than 50 papers in referred National and International conferences/Journals. His area of interest includes Digital Communication, Soft Computing, Robotics and etc.

A Feedback Design for Rotation Invariant Feature Extraction in Implementation with Iris Credentials

M. Sankari

Department of Computer Applications,
Nehru Institute of Engineering and Technology,
Coimbatore, INDIA.

R. Bremananth

School of EEE, Information Engg. (Div.),
Nanyang Technological University,
Singapore.

Abstract—Rotation invariant feature extraction is an essential objective task in computer vision and pattern credentials problems, that is, recognizing an object must be invariant in scale, translation and orientation of its patterns. In the iris recognition, the system should represent the iris patterns, which is invariant to the size of the iris in the image. This depends upon the distance from the sensors to subjects' eye positions and the external illumination of the environments, which in turn make the changes in the pupil diameter. Another invariant factor is the translation, the explicit iris features should be a positional independent even though eye present anywhere in the acquired image. These two invariants are perfectly achieved by the weight based localization approaches. However, the iris orientation estimation is an important problem to avoid in preserving selective orientation parameters. Multiple source points are used to estimate the segmented objects orientations. After estimating the deviation in angle of segmented object that can be rotated to its principal origin and then the feature extraction process is applied. A multi resolution approach such as wavelet transform is employed for feature extraction process that provides efficient frequency and spatial texture feature deviations present in the irises. In this paper, we work on a feedback design with Radon transform with wavelet statistical analysis of iris recognition in two different ways. In order to check the viability of the proposed approaches invariant features are directly compared with weighted distance (WD) measures, in the first phase and second phase is to train the Hamming neural network to recognize the known patterns.

Keywords- *Iris credentials; Invariant Features; Rotation estimation; Multiresolution analysis;*

I. INTRODUCTION

In computer vision and pattern recognition, rotation invariant feature extraction is an essential task, that is, recognizing an object must be invariant in scale, translation and orientation of its patterns. This paper emphasizes on invariant feature extraction and statistical analyses. In the iris recognition, the system should represent the iris patterns, which is invariant to the size of the iris in the image. This depends upon the distance from the sensors to subjects' eye positions and the external illumination of the environments that make the changes in the pupil diameter. Another invariant factor is the translation where iris features should be a positional independent of iris pattern, it could occur anywhere in the

acquired eye image. However, the iris orientation estimation is an important problem to avoid in preserving selective orientation parameters, for example, 7 relative orientations were maintained for iris best matching process in the literature [1] and seven rotation angles (-9, -6, -3, 0, 3, 6 and 9 degrees) used by Li ma et al. [2]. In the real time imaging, due to the head tilt, mirror angle and sensor positions, iris images are captured in widely varied angles or divergent positions. We estimate the rotation angle of iris portion within the acquired image by using multiple line integral approaches, which provide better accuracy in the real time capturing. Local binary patterns, gray-level and auto-correlation features were used to estimate orientation of the texture patterns. It projected the angles that are locally invariant to rotation [3]. In [4], texture rotation-invariant was achieved by autoregressive models. It used several circle's neighborhood points to project the rotation angle of the object. Aditya Vailaya et al. [5] had dealt with Bayesian learning framework with small code features that are extracted from linear vector quantization. Thus, these features can be used for automatic image rotation detection. A hidden Markov model and multichannel sub-band were used for estimating rotation angles of gray level images in the study [6]. In this work, we propose Radon transform based multipoint sources to estimate the rotation angle estimation for real-time objects.

Classification is a final stage of pattern recognition system where each unknown pattern is classified to a particular category. In iris recognition system, a person is automatically recognized based on his / her iris pattern already trained by the system. This is done in a way of training a brain to teach certain kind of sample patterns. In the testing process, system recalls the trained iris patterns as a weighted distance specified by the system. If threshold is attained then system genuinely accepts a person, otherwise false alarm sounds. However, the way to find the statistical level is a tedious work because it makes decision to evaluate the pattern either genuine or fake. Hence combinatorics of iris code sequence should be carried out by means of statistical independence. Moreover, failure of iris recognition is principally concerned with a test of statistical independence because it absorbs more degree-of-freedom. The test is nearly assured to be allowed whenever the extracted iris code comparing from two different eyes are evaluated. In addition, the test may exclusively fail when any iris code is compared with another version of itself. The test of statistical independence was implemented by the Hamming distance in

[1] with a set of mask bits to prevent non-iris artifacts. Li ma et al. [7] proposed a classifier design that was based on exclusive-OR operation to compute the match between pairs of iris bits. In [2], authors worked with the nearest centre classifier to recognize diverse pair of iris patterns. A competitive neural network with linear vector quantization was reported for both identification and recognition of iris patterns by Shinyoung Lim et al. [8]. Our main contribution to this paper is a feedback design (Fig. 1) to extract an appropriate set of rotation invariant features based on Radon and wavelet transforms. An iteration process is repeated until a set of essential invariant features is extracted from the subject. We have done two different phases of statistical analyses of rotation invariant iris recognition. During phase I, wavelet features are directly compared with weighted distance (WD) measures and in phase II invariant features were trained and recognized by the Hamming neural network.

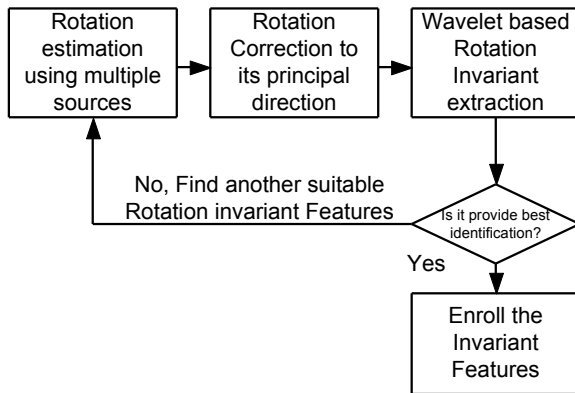


Fig. 1. A feedback design of rotation invariant feature extraction.

The remainder of this paper is organized as follows: Section II emphasizes on invariance and estimation of rotation angle. Radon and wavelet based rotation invariant is described in section III. Section IV depicts the results obtained based on the proposed methodologies while Concluding remarks and future research direction are accentuated Section V.

II. INVARIANCE IN ROTATION

A 2D rotation is applied to an object by repositioning it along a circular path. A rotation angle θ and pivot point about which the object to be rotated is specified for generating series of rotation. In counterclockwise, positive angle values are used for rotation about the pivot point and in contrast clockwise rotation requires negative angle values. The rotation transformation is also described as a rotation about an axis that is perpendicular to the xy plane and passes through the pivot point. The rotation transformation equations are determined from position (x_1, y_1) to position (x_2, y_2) through an angle B relative to the coordinate origin. The original displacement of the point from the x-axis is, angle A. By trigonometric ratios, $\sin(A) = y_1 / r$, $\sin(A + B) = y_2 / r$, $\cos(A + B) = x_2 / r$ and $\cos(A) = x_1 / r$. From the compound angle formulae described as

$$\sin(A + B) = \sin(A) \cdot \cos(B) + \cos(A) \cdot \sin(B). \quad (1)$$

Substituting trigonometric ratios and obtain the following

$$(y_2 / r) = (y_1 / r) \cdot \cos(B) + (x_1 / r) \cdot \sin(B), \quad (2)$$

$$y_2 = y_1 \cos(B) + x_1 \sin(B), \quad (3)$$

$$y_2 = x_1 \sin(B) + y_1 \cos(B), \quad (4)$$

Likewise, substituting trigonometric ratios and derived as

$$\cos(A + B) = \cos(A) \cdot \cos(B) - \sin(A) \cdot \sin(B), \quad (5)$$

$$(x_2 / r) = (x_1 / r) \cdot \cos(B) - (y_1 / r) \cdot \sin(B), \quad (6)$$

$$x_2 = x_1 \cos(B) - y_1 \sin(B), \quad (7)$$

Therefore, from Eqs. (5) and (10) we can get counterclockwise rotation matrix and the new coordinate position can be found as described in Eq. (8). The basics of rotation and line integrals are incorporated together to form equations for projecting the object in single and multi source points.

A. Multipoint source

Based on the basics of rotation, multipoint source method computes the line integrals along parallel beams in a specific direction. A projection of image $f(x,y)$ is a set of line integrals to represent an image. This phase takes multiple parallel-beams from different angles by rotating the source around the centre of the image.

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos(B) & -\sin(B) \\ \sin(B) & \cos(B) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}. \quad (8)$$

This method is based on Radon transform, which estimates the angle of rotation using the projection data in different orientations. A fusion of Radon transform and Fourier transform had been performed for digital watermarking which is invariant to the rotation, scale and translation invariant in the literature [9]. A parallel algorithm for Fast Radon transform and its inverse was proposed by Mitra et al. [10]. Radon transform was employed for estimating angle of rotated texture by Kourosh et al. [11]. Image object recognition based Radon transform was proposed by Jun Zhang et al. [12], this method is robust and invariant to rotation, scale and translation of image object. Fig. 2 shows a multipoint source at a specified angle for estimating rotation angle of a part of iris. This method projects the image intensity with a radial line orientation at a specific angle from the multipoint sources. Multipoint projection computes any angle θ by using the Radon transform $R(x', \theta)$ of $f(x,y)$, it is the line integral of parallel paths to the y axis. After applying the function of multipoint sources $R(x', \theta)$ the resultant data contain row and column. Column describes projection data for each angle in θ and it contains the respective coordinates along the x' axis. The procedure for applying multipoint source projection to estimate the angle is as follows: Image is rotated to a specific angle in counterclockwise by bi-cubic interpolation method.

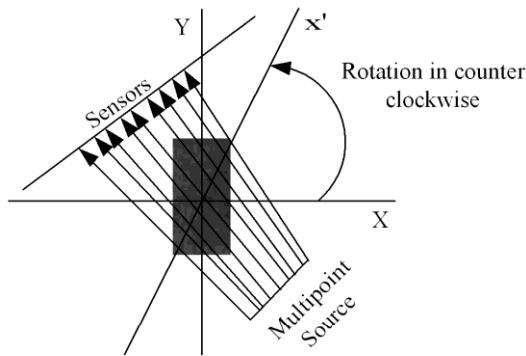


Fig. 2. Multipoint estimation using multipoint sources.

Assume the rotation angle from 1° to 180° in order to find the peak area of rotation angles. After applying the multipoint sources, Radon transform coefficients have been generated for each angle. The standard deviation of the Radon transform coefficients is calculated to find the maximum deviation of rotation angle. This is shown in Fig. 3. Then, using estimated angle, object rotation is rotated to its original principal angle using bi-cubic interpolation method. If the estimated angle $\hat{\theta}$ is positive then rotate the object as $-(\hat{\theta} + 90^\circ)$ in clockwise direction else if the estimated angle is negative or above 90° then rotate the object as $-(\hat{\theta} - 90^\circ)$ in clockwise direction.

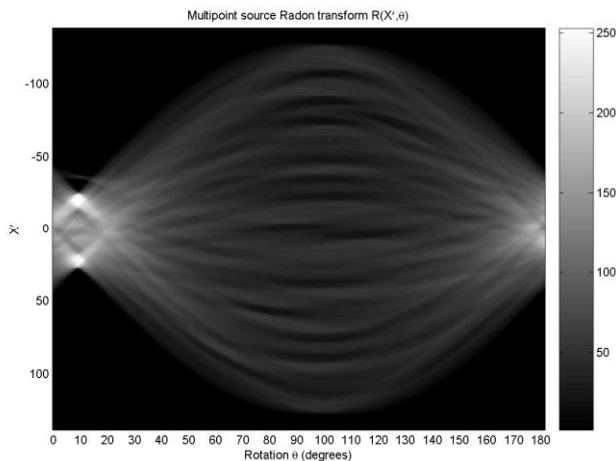


Fig. 3. Illustration of orientation angle estimation using multipoint.

III. IRIS WAVELET FEATURE ANALYSIS

In this phase wavelet based feature extraction process has been employed to extract feature obscured in the iris patterns. It is an essential task for recognising a pattern from others because some features may produce same type of responses for diverse patterns. It causes the hypothesis in pattern recognition process to differentiate one from another. To overcome the problem of uncertainty the system needs an efficient way to extort quality features from the acquired pattern. Iris provides sufficient amount of interclass variability and minimises intra-class variability. Thus the characteristics of these patterns are well efficiently taken out by the sense of using less computational process. Among various feature extractors,

wavelet series approximate hasty transitions much more accurately than Fourier series. Consequently, wavelet analysis perfectly replicates constant measurements. It produces better approximation for data that exhibit local variation and because of its basis function each term in a wavelet series has a compact support within a finite interval. The other sense to employ wavelet is orthogonal. This means that information carried by one idiom is independent of information conceded by the other. Thus, there is no redundancy in the feature extraction. This is fine when neither computational sequence time nor storage is wasted as a result of wavelet coefficient computed or stored. The next sense related with wavelet is multi resolution, which is like biological sensory system. Many physical systems are organised into divergent levels or scales of some variables. It provides an economic structure and positional notion of arithmetic whose computational complexity is $O(N)$, where N data points are to be accessed [13]. In the current literature various computer vision and signal processing applications have been based on wavelet theory [14] such as detecting self-similarity, de-noising, compression, analysis and recognition. This technique has proven the ability to provide high coding efficiency, spatial and quality features. However, wavelets features are not rotation invariant due to its directional changes. Hence this approach initially estimates the extorted pattern rotation angle and rotates to its principal direction. Afterwards multi resolution wavelets have been employed to extort features from the rotation corrected patterns. In the iris recognition process, this approach has adopted Daubechies (db) wavelet to decompose the iris patterns into multiple resolution sub-bands. These sub-bands are employed to transform well-distributed complex iris patterns into a set of one-dimensional iris feature code. Decay is a process to divide the given iris image into four sub-bands such as approximation, horizontal, vertical, and diagonal coefficients. A 2D Daubechies wavelet transform of an iris image (I) can be carried by performing two steps, Initially, it performs 1D wavelet transform, on each row of (I) thereby producing a new image I_1 . In second step it takes I_1 as an input image and performs 1D transform on each of its columns. A Level-1 wavelet transform of an image can be described as

$$I \Rightarrow \begin{bmatrix} a^1 & h^1 \\ v^1 & d^1 \end{bmatrix}, a^1 \Rightarrow \begin{bmatrix} a^2 & h^2 \\ v^2 & d^2 \end{bmatrix}, \quad (9)$$

where the sub-images a^1 , h^1 , v^1 and d^1 represent level-1 approximation, horizontal, vertical and diagonal coefficients a^2 , h^2 , v^2 and d^2 level 2 coefficients. The approximation is created by computing trends along rows of I followed by computing trends along columns. Trends represent the running average of the sub-signals in the given image. It produces a lower frequency of the image I . The other sub-signals such as horizontal, vertical and diagonal have been created by taking fluctuation. It is a running difference of sub-signals. Each coefficient represents a spatial area corresponding to one-quarter of the segmented iris image size. The low and high frequencies represent a bandwidth corresponding to

$0 < |\omega| < \pi/2$ and $\pi/2 < |\omega| < \pi$, respectively. Fig. 4 shows frequency variation of Daubechies wavelets. The wavelet transform is defined as

$$W(a, \tau_x, \tau_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I(x, y) \psi_{a, \tau_x \tau_y}(x, y) dx dy, \quad (10)$$

$$\psi_{a, \tau_x \tau_y}(x, y) = \frac{1}{|a|} \psi\left(\frac{x - \tau_x}{a}, \frac{y - \tau_y}{a}\right), \quad (11)$$

where $I(x, y)$ is a segmented iris image, $W(a, \tau_x, \tau_y)$ is a wavelet transform function, $\psi_{a, \tau_x \tau_y}(x, y)$ the wavelet basis function, a is a scaling factor, τ_x and τ_y are translation factors of x and y axes, respectively. The properties separability, scalability, translatability of discrete wavelet transform is performed as

$$\phi(x, y) = \phi(x)\phi(y), \psi^H(x, y) = \psi(x)\phi(y), \quad (12)$$

$$\psi^V = \phi(x)\psi(y), \psi^D = \psi(x)\psi(y), \quad (13)$$

$$W\phi(l_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=1}^M \sum_{y=1}^N I(x, y) \phi_{l_0, m, n}(x, y), \quad (14)$$

$$W\psi^H(l, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=1}^M \sum_{y=1}^N I(x, y) \psi_{l, m, n}^H(x, y), \quad (15)$$

$$W\psi^V(l, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=1}^M \sum_{y=1}^N I(x, y) \psi_{l, m, n}^V(x, y), \quad (16)$$

$$W\psi^D(l, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=1}^M \sum_{y=1}^N I(x, y) \psi_{l, m, n}^D(x, y), \quad (17)$$

where $\phi(x, y)$ and $W\phi(l_0, m, n)$ are scaling function and approximation coefficients of $I(x, y)$ at scale l_0 , respectively.

$W\psi^H(l, m, n)$ $W\psi^V(l, m, n)$ $W\psi^D(l, m, n)$ are coefficients of horizontal, vertical and diagonal details for scales $l \geq l_0$ respectively.

Normally $l_0 = 0$, and assigning $M = N = 2^L$ so that $l = 0, 1, 2, \dots, L-1$ and $m = n = 0, 1, 2, \dots, 2^j - 1$.

The decomposition of signals produces sub-signals such as low, middle and high frequency of the components, which play a very important role in the feature extraction process. In this approach Daubechies wavelet is employed for feature extraction process. Its frequency distribution for different level is illustrated in Fig. 5. The DWT (Discrete Wavelet Transform) consists of $\log_2 N$ stages if the given signal s is of length N . Fig. 6 shows the scaling and wavelet functions of Daubechies wavelets. Initially s produces two sets of coefficients such as approximation coefficients cA_1 , and detail

coefficients cD_1 . These coefficients are obtained by convolving s with the low-pass filter Lo_D for approximation, and with the high-pass filter Hi_D for detail coefficients. In the case of images, a similar procedure is possible for 2D wavelets and scaling functions obtained from one-dimensional wavelets by tensorial products. This kind of 2D DWT leads to a decomposition of approximation coefficients at level j in four components: the approximation at level $j + 1$ and the details in three orientations (horizontal, vertical, and diagonal). Fig. 7 shows the decomposition process.

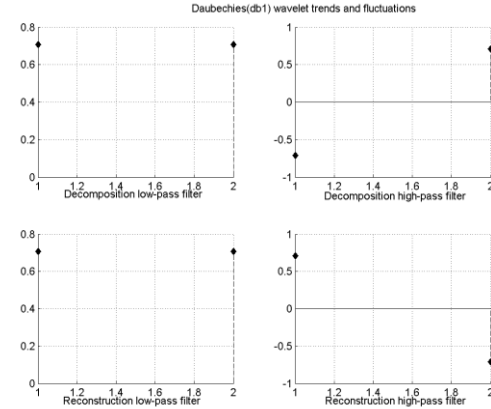


Fig. 4. Daubechies (db1) wavelets frequency variations.

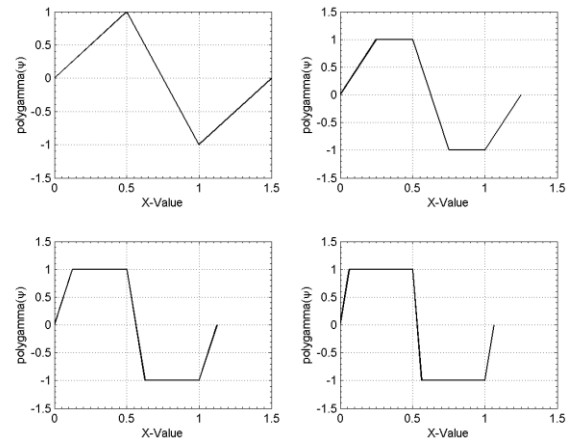


Fig. 5. Frequency distribution of Daubechies wavelets by different iterations.

In the feature extraction process of iris patterns four levels of decompositions have been performed to obtain fine level of frequency details from the pattern. The scaling factor is very important for decomposing the given iris signals. At the first level it produces 648 signals, second level has 162 signals, third level 45 signals and finally it generates 15 signals for each frequency. The MRA produces the frequency signals to compact approximation of features which aid to generate an efficient set of distinct features that are provided with less intra class variability and more interclass variability in the iris pattern recognition process. Fig. 8 shows four levels of decomposition process for the given iris images. Low-pass

filter corresponds to an averaging operation and extract the coarse information of the signals, whereas high-pass filter corresponds to a dissimilarity operation that extracts the detailed information of the signals.

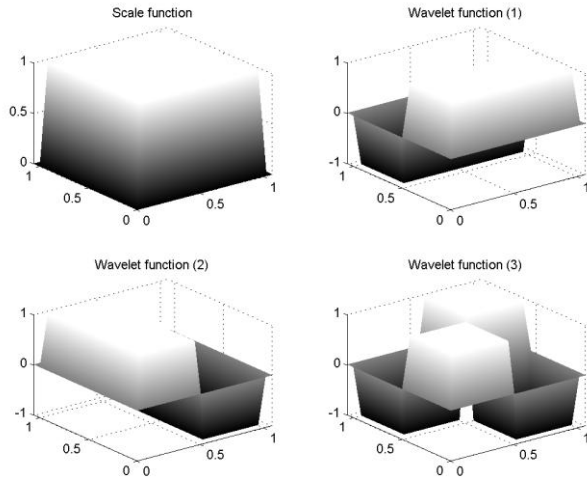


Fig. 6. Scaling and wavelet functions of Daubechies wavelets.

When iris signal passes through these low and high pass filters, it generates the frequency variation occurring in a pattern.

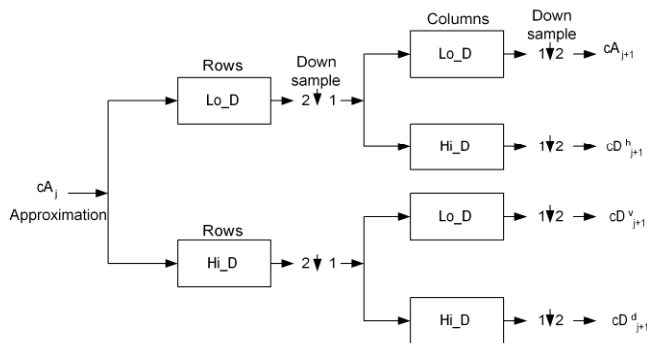


Fig. 7. Decomposition of wavelet signals in the feature extraction.

A. Iris feature selection

In this phase frequency variation of iris signals in divergent levels are quantized into iris features. For that multi resolution frequencies of low and high pass filters are taken for quantization process of conversion of real signals into binary. The mean and standard deviation of approximation and detail coefficients vary in each level of the decomposition of iris patterns which raises up to generate an efficient feature sets of the given patterns. The horizontal, vertical and diagonal coefficients wavelet features have middle and high frequencies of the components of iris signals. The histogram analyses of signals in divergent levels are illustrated in Fig. 9. The frequency distribution of signals at level 1 ranges from -10 to 10 and from -100 to 100 at level 4 for horizontal coefficients.

Thus this approach quantizes these trends and fluctuation of sub-signals into iris features. After performing the four level of decay process the horizontal, vertical and diagonal coefficients, the iris are used for iris feature encoding process. The frequency variation occurring in these decomposition coefficients are employed to extract iris feature codes. In order to make an efficient set of features and reduce the computational time of iris matching process the coefficient values are converted into binary values which senses to create a compact feature set.

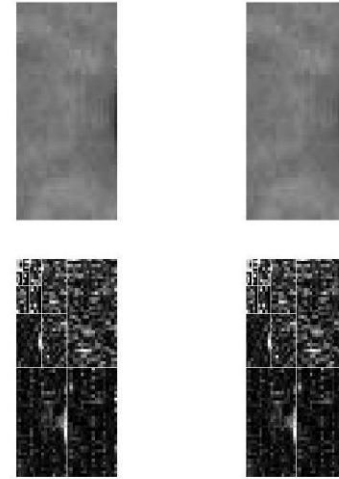


Fig. 8. Four level of decomposition of iris patterns.

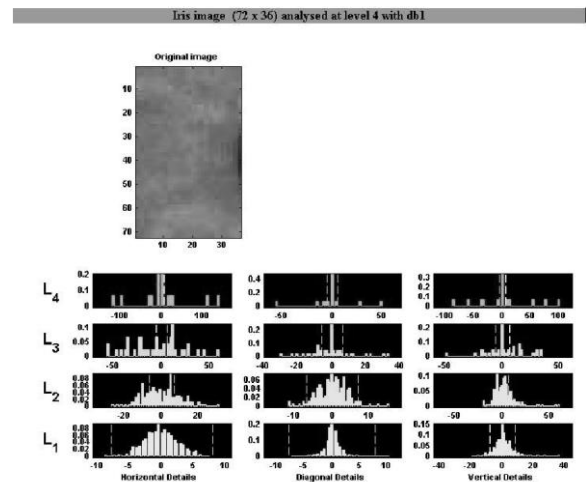


Fig. 9. Histogram of divergent levels of iris image.

In the current literature, Haar wavelets are used for iris image feature extraction by decomposing the signals into four levels [8]. It uses only high frequency of the components for representing iris patterns. However, iris patterns are having middle frequency of the components, which are essential for recognizing iris patterns in large population. Moreover, in their approach there is no transformation-invariant analysis. When there is a rotation between a set of irises from the same subject, it may produce false positives in the recognition

because these patterns produce different kinds of features for diverse rotation and translations. Here, transformation invariant analysis is performed before extracting features from the iris patterns. In addition, middle and high frequency of iris wavelet features were extracted for recognition. Thus, it reduces less false positives in the recognition process using a feedback design based on the rotation invariant features. Though iris patterns are unique in nature, it is a difficult process to generate identical template for the same subject. This is mainly due to the changes in imaging position, distance, illumination conditions, eyelashes / eyelid occlusion and eyewear reflections. These factors may affect the efficacy of the system. Thus this approach compensates the deformation of these factors and recognizes the iris patterns, which are independent of transformation factors and other artifacts. The classification results of rotation invariant and wavelet features are illustrated in Section V.

VI. ROTATION INVARIANT CLASSIFICATION

The different pairs of eye images were captured in diverse distances and illuminations provide more challenges to this approach. Experimentations were also performed with different eye images in diverse criteria like normal, outdoors, contact lens, spectacles, and diseased (Tumours, Tear, Iridocyclities) eyes. The database of iris images has 2500 images captured from 500 different subjects as each has been acquired as 5 different images with different real-time conditions [18, 19]. In the iris matching process, inter and intra class iris features are efficiently separated and they prevent impostors from entering into the secure system. To authenticate any genuine user, iris feature sets are treated as trained sets and stored in the encrypted file. Verification subjects' irises are represented as test sets. The same subject iris feature codes could vary due to external noises, lighting, illuminations and other factors such as closed eyelashes or eyelids. This possibly will lead to different iris template for an eye, even though iris is unique in nature. However, capturing eye images with advanced biometric camera may solve this problem. The process by which a user's biometric data is initially acquired, validated, processed and stored in the form of a template for ongoing use in a biometric system is called enrolment. Quality enrolment is a critical factor in the long-term accuracy of biometric system. Wavelet features of irises are recognized using the weighted distance (WD). It recognizes the various classes of iris codes by checking a minimum distance between two iris codes. This is defined as, $\min_i WD(IFC(x_{trained}), IFC(x_{test}))$, where $WD(IFC(x_i), IFC(x_j))$ represents weighted distance in between two iris feature sets as defined as

$$WD(IFC(x_{trained}), IFC(x_{test})) = \frac{|IFC(x_{trained}) \oplus IFC(x_{test})|}{N}, \quad (18)$$

where N denotes the number of bits in the iris feature set. The weighted distance (WD) is used to determine the number of error bits in between two iris classes. In the experimentation,

the weighted distance of the intra class feature set is discriminated by the constraint $0 \leq WD \leq 0.2$ and inter class iris features is abandoned with the constraint $WD > 0.2$. These distances are also evaluated based on the normal distribution of mean, standard deviation and degree-of-freedom of the wavelet iris codes. In addition the same candidate's iris image may have more artifacts due to various deteriorates as stated previously. Hence WD needs more discriminability range for recognising the genuine subject. Conversely, if system maintained large distance variation to allow the subjects, then more FAR (False accept rate) might be encountered. Moreover, if WD is reduced then more FRR (False rejection rate) may be produced by the system. The system was tested with normal and abnormal images and their mean of weighted distance of genuine-class iris codes was $\mu=0.10813$, its standard deviation was $\sigma=0.0392$ and degree-of-freedom v was 62.621991. Impostor-class mean value was $\mu=0.27104$ and its standard deviation was $\sigma=0.040730$. During the weighted distance computation, an identical iris pattern was produced $WD = 0$ and due to abnormal conditions the same subject iris was assorted from 0 to 0.19 WD. This is shown in Fig. 10. If distributions are very large then system allows more changes for impostors to access the system. This type of limitation of distributions may be provided with more false reject rate, but minimum false accept rates. In most of the applications such as Bank-ATM and biometric voting machines these type of constrained weighted distance are essentially desirable in order to agree entire genuine subjects. In the recognition phase, GAR (Genuine accept rate) was 99.3% and FAR was 0.7% and in confirmation MR (Matching rate) was 99.94% and FRR was 0.06%.

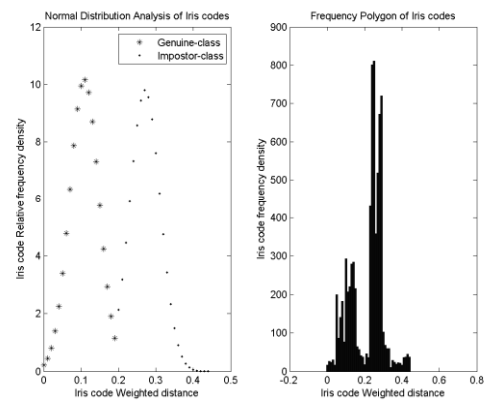


Fig. 10. Weighted distance distribution for wavelet iris features and frequency polygon of the iris codes.

A. Hamming neural network (HNN)

Hamming neural network (HNN) is an alternative way to train and test the extracted features [15]. This network is employed to train for both iris and character patterns. Its input layer can accept wavelet features. That is, it works with bipolar value of the extracted iris wavelet features. Wavelet based iris feature codes are fed for recognition process. HNN is used to recognize iris features from the trained set. The aim of the

HNN is to decide which trained iris feature set is closest to the test feature set [16]. HNN consists of two layers; the first layer is called as a feed forward layer (FFL) that is used to calculate a maximum score of the input patterns and a recurrent layer (RL) is used to select the maximum score among the input patterns. Each neuron in the FFL is set up to give maximum response to one of the trained patterns. If test set is same as trained set the maximum score is taken by the recurrent network. The weights initialization process of the HNN is described as

$$W_{ij} = \frac{x_{ij}}{2}, W_{i0} = n/2, \quad (19)$$

where w_{ij} and x_{ij} are the weights value and input features of j^{th} bit of the i^{th} iris feature, w_{i0} is a bias value and n is the number of bits in the iris features. In order to incorporate HNN with iris recognition, wavelet features are converted from binary to its corresponding bipolar form. For example, the i^{th} iris feature set is $\{-1, +1, +1, -1, +1 \dots +1\}$. The weight of i^{th} neuron is set to $\{w_{i0} = 67.5, w_{i1} = -0.5, w_{i2} = 0.5, w_{i3} = 0.5, w_{i4} = -0.5, w_{i5} = 0.5, \dots, w_{i135} = 0.5\}$. The weighted sum is 135. Each of the neurons in the FFL gives a maximum response of 135 for exactly identical iris codes, and a minimum value to other features. In HNN, the number of neurons in the FFL is same as the number of neurons in the recurrent layer. When a test feature is given to the FFL, the output from each of the neurons in the FFL is measured by the Hamming distance from the iris in the training set. The Hamming distance between two iris patterns is a measure of the number of bits that are different between the two iris patterns. For example, if an input iris pattern of $\{+1, +1, +1, -1, -1 \dots +1\}$ is fed then the output of FFL is 133 which has 2 less than the maximum of 135. This is because the given pattern has 2-bit difference. Perhaps, if entire bits are changed in the iris patterns, the neuron that corresponds to that pattern produces an output of 0. The function of RL is to select the neuron with the maximum output. The final output of the RL contains a large positive value at the neuron that corresponds to the nearest iris pattern, and all other neurons produce 0 value. The RL is trained by setting the weight to 1, if the weight connection corresponds to the same neuron and all other weight value are small negative value less than $-1/TI$. The response of the RL is described as

$$Y = \begin{cases} \sum_{i=0}^{\pi} w_i \cdot x_i & \text{if } \sum_{i=0}^{\pi} w_i \cdot x_i \geq \lambda \\ 0 & \text{otherwise} \end{cases}, \quad (20)$$

where TI is the total iris patterns available in the trained set, λ is a threshold maintained in the iris recognition process. In this process, the output is fixed to the value of the output of the FFL. The RL is allowed to iterate, initially the outputs of the RL is equal to the score produced by the FFL. Then, because of the less than $-1/TI$ weights, the output is gradually reduced. After some iteration, all the outputs reach 0 except the recognized pattern with threshold, for example, $\lambda_h = 81$, i.e. weighted distance for HNN, WD_h is 0.6. The testing process

of HNN is stopped if there is no change expected in the iterations.

V. EXPERIMENTS INVARIANT CLASSIFICATION

Experiments were carried out on cases like left and right eye evaluation, twins eye evaluation, eyewear and artifact evaluation, hypothesis test, segmented iris, normalized iris, Receiver operating characteristics curve (ROC) evaluations and feature vector dimension variations. To evaluate these phases, system was tested based on GAR, FAR and FRR factors of the recognition.

A. Fusion of left and right eyes

Evaluating both the left and right eye combinations provide better security in the application domains. However, the recognition time is directly prepositional with the number of entries in the iris database. A pair of 120 subjects' eye images was acquired to test the algorithm, that is, a total of 240 iris patterns were trained and tested by the ENDM, weighted distance and HNN. The feature vector size is double the dimension of normal vector. Thus, 270 wavelet iris features were computed for each subject to test the weighted distance and HNN. Table I depicts the recognition rates for evaluating both left and right eye images. In the recognition process, a system was set by a matching threshold level. It determines the error tolerance of the system with respect to the features for which the network is trained, and is used to determine whether the final result is accepted / rejected. For any recognition system that is used for security applications, this error tolerance should be minimal and therefore the setting of this matching threshold is a crucial factor. Recognition rates were reported based HNN and WD. WD was better recognition rates with minimal FAR. Furthermore, its FRR was also an acceptable one, hence the system with wavelet and WD produce good performance than the HNN.

TABLE I. COMPARISON OF CLASSIFIERS ACCURACY RATE

Feature	Types of classifier	Left and right iris features	Matching Threshold Range	Recognition rate		
				GAR %	FAR %	FRR %
Wavelet	WD	270	[0.0-0.19]	99.4	0.6	1.2
Wavelet	HNN	270	[0.7-1.0]	99.32	0.68	2.7

B. Recognition of twins

Identical twins' irises were separately verified with different methods. From 50 twins, 500 eye images were acquired. It contained both left, right eye images with each subject having 10 eye images. The twins' iris code result was generated by the classifiers as the same weighted distances as the regular iris codes available in the database. The mean of WD in the images acquired from twins is 0.086360 with the standard deviation 0.044329. A confidence interval is a range of values that have a chosen probability of containing the true

hypothesized quantity. The standard deviation of confidence intervals is in the range 0.0417 to 0.0473. Fig. 11 shows the normal distribution of twin's iris code weighted distance. In the checking of twin's iris database WD was changed in the range 0 to 0.19, GAR was 99.3 and FAR was 0.7.

C. Eyewear and Artefacts

Eyewear images are major problematic ones in the iris recognition because these images may produce more false localization and FAR or FRR of the system. To evaluate the recognition rates of eyewear images, 50 subject's eye images were acquired with white glasses from each of them 5 images were captured, that is total of 250 images were utilized for the recognition. As stated previously, an exact identical iris pattern could be produced $WD=0$, but due to eyewear noises and other artifacts its patterns require a certain WD range. Hence system evaluated the same subjects' iris patterns before and after wearing the eyewear. This also included soft contact lens and white glass with different varieties. Table II shows the WD on the image with and without wearing eyewear. In that hard contact lens produced more FRR in the recognition. Thus it was around 32 bits average error bits and WD was 0.237. Moreover, localization system may be disrupted by the designed frames of eyewear in the hypothesis to locate the ROI. However, in the recognition, it produced minimal average error of 22-bits.

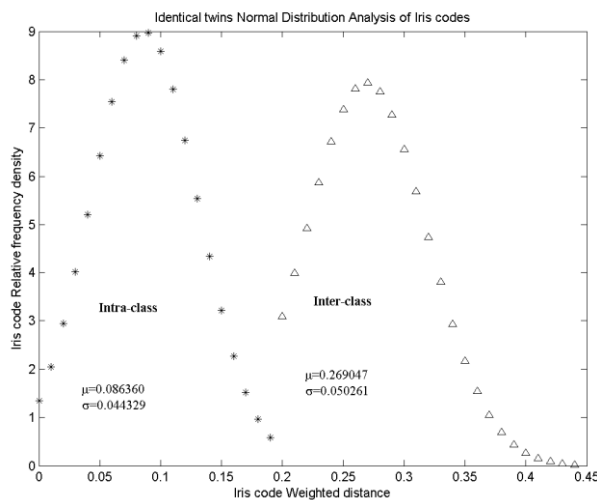


Fig. 11. Representation of weighted distance of twins' iris code.

The FRR and FAR was high when images were acquired with eyewear and in diverse illuminations such as sunshine and twilight conditions, eyewear at twilight the average of 34 bits were corrupted, therefore WD was 0.251. As a consequence, the system recommends the application domain while enrolling a void eyewear because during enrolment iris patterns could be signed up with minimum amount of error bits. Therefore, it increases system recognition rates in order to achieve better rotation invariant feature set.

D. Iris hypothesis test

Hypothesis test plays very important role in the biometric recognition system, i.e., making a decision is based on the availability of data during the training or enrolment and testing or verification processes.

TABLE II. EYEWEAR NOISES AND OTHER ARTIFACTS ASSESSMENT

Types of eye wears	Without eyewear		With eyewear	
	Average Error bits out of 135	Average error in WD	Average Error bits out of 135	Average error in WD
White glass	14	0.104	22	0.165
Soft contact lens	15	0.112	26	0.194
Hard contact lens	15	0.112	32	0.237
Sunshine	18	0.129	27	0.198
Twilight	19	0.138	34	0.251

This test may be neither true nor false. It could be dependent on the feature extraction and classifier design of the system. Thus this system makes iris images as transformation invariant patterns to increase the performance of the system. The biometric estimation is based on some terms of assumptions that is, make a system as null hypothesis. The null hypothesis is the original declaration. In iris recognition the null hypothesis is specified by the WD range between 0.0 and 0.2 for the genuine subject. The significance level is another term related to the degree of certainty that the system requires in order to reject the null hypothesis in favor of the alternative hypothesis. By taking a small sample the system cannot be certain about the conclusion. So decide in advance to reject the null hypothesis if the probability of observing the sampled result is less than the significance level. A typical significance level is 0.21. The p-value is the probability of observing the given sample result under the assumption that the null hypothesis is true. If the p-value is greater than the WD range, then system rejects the null hypothesis. For example, if $WD=0.2$ and the p-value is 0.22, then the system rejects the null hypothesis. The results of biometric for many hypothesis tests also include confidence intervals. That is, a confidence interval is a range of values that have a chosen probability of containing the true hypothesized quantity. An illustrative example, $WD=0.03$ is inside a 97% confidence interval for the mean. That is equivalent to being unable to reject the null hypothesis at a significance level of 0.03. Conversely, if the $100(1-WD)$ is confidence interval that does not contain weighted distance range then the system rejects the null hypothesis at the level of significance.

E. Receiver operating characteristics curve analysis

The ROC analysis of wavelet features with WD and HNN is illustrated in Fig 12. The both WD and HNN classifiers produced approximately the same amount of accuracy. However, WD produced quite better exactness than the HNN since it requires minimal error tolerance and threshold in

training and testing processes. Thus HNN loosely allows impostors more than WD.

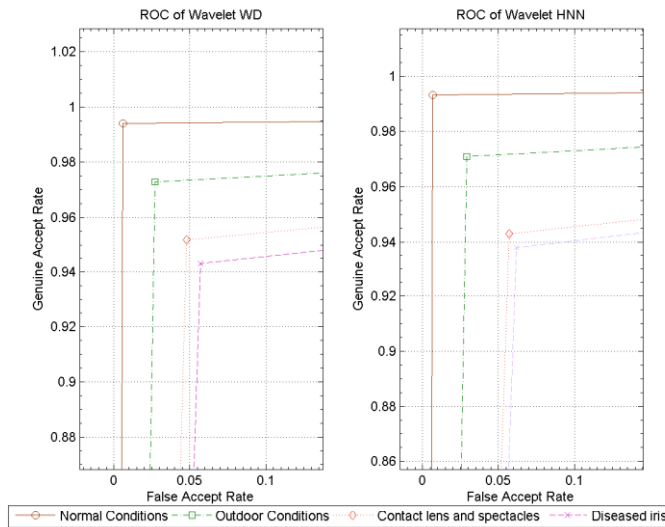


Fig. 12. ROC analysis wavelet features with WD and HNN.

F. Performance comparison

In this work, a feedback design for extraction of rotation invariant iris recognition based on local segmentation of iris portions was suggested. It prevents misclassifications (FAR) of iris patterns and limits the overall FRR of the system. As per research work, 40% of iris images have been obscured by eyelids / eyelashes and 35% of images hid the top portions of iris. This system pulls out left, right and bottom local area of iris for iris code extraction. It provides overall accuracy of 98.3% in the iris localization process. In [2], elastic deformation has occurred in the iris portion due to illumination changes. It was compensated to convert the circular portion of the iris (including eyelids / eyelashes) into a rectangle strip, which was used for convolution operations in iris matching. In the present work these types of discrepancies have been resolved by local segmentation process. In addition, the previous method influenced by seven rotation angles (-9, -6, -3, 0, 3, 6 and 9 degrees). But in our proposed system, the rotation-invariance was achieved by on combination of Radon transform and wavelet feature sets. In [1], 2048 feature components were used to classify the diverse iris patterns. It readily achieved scale and translation invariant pattern analysis using integrodifferential operator. However, rotation-invariant might be carried out by shifting of iris phase codes. So, it inclined sequences of orientation of templates for the recognition process. In our approach, we employed with sequences of rotation estimation preprocessing based on the Radon transform in order to extract the rotation invariant features, which in turn, influence a distinctive template for each subject in enrolled of the system. In [3], Shinyoung Lim et al. suggested an approach based on Haar wavelet with linear vector quantization method. This method worked with 87 high pass filter of the wavelet transformation. However, middle frequencies of the iris patterns are very useful in the recognition. In our present work both middle and high

frequencies of wavelet component of iris patterns are used. Additionally, transformation-invariant is efficiently achieved prior to the feature extraction. Therefore, multiple iris features or additional shift operation is completely avoided in the proposed methodology. Thus, this paper provides better accuracy with compact rotation invariant feature set than previous methods.

IV. CONCLUSION AND FUTURE WORK

This paper processes a feedback design for rotation invariant feature extraction in application with iris patterns using Radon and wavelet analysis. After correcting rotation angle, rotation invariant contours are processed by feature extractor repeatedly until a suitable set was encountered. It increases more recognition rate and rotation estimation with diverse artifacts than the other methods since the previous methods used redundant patterns of iris feature templates for different angle of capturing or additional shift operation for compensating the invariants. Suggested methods would be possibly implemented with other applications of object rotation estimation and recognition. This paper opens a new direction of research in the vision and biometric committees.

ACKNOWLEDGMENT

Authors thank their family members and children for their continuous support and consent encouragement to do this research work successfully.

REFERENCES

- [1] Daugman J., 'How Iris Recognition Works', IEEE Transactions On Circuits and Systems For Video Technology, vol. 14, no. 1, pp. 21-30, 2004.
- [2] Li ma, Tieniu Tan, Yunhong Wang and Dexin Zhang, 'Personal Identification Based on Iris Texture Analysis', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp. 1519-1533, 2003.
- [3] Pietikainen M., Ojala T. and Xu Z., 'Rotation-invariant texture classification using feature distributions', Pattern Recognition, vol. 33, pp. 43-52, 2000.
- [4] Mao J. and Jain A. K., 'Texture classification and segmentation using multiresolution simultaneous autoregressive models', Pattern Recognition, vol. 25, no. 4, pp. 173-188, 1992.
- [5] Aditya Vailaya, Hong Jiang Zhang, Changjiang Yang, Feng-I Liu and Anil K. Jain, 'Automatic Image Orientation Detection', IEEE Transactions on Image Processing, vol. 11, no. 7, pp. 746-755, 2002.
- [6] Chen J. L. and Kundu A. A., 'Rotation and Gray scale transformation Invariant Texture Identification Using Wavelet Decomposition and Hidden Markov Model', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, no. 2, pp. 208-214, 1994.
- [7] Li ma, Tieniu Tan Yunhong Wang and Dexin Zhang, 'Efficient Iris Recognition by Characterizing key Local variations', IEEE Transaction on Image Processing, vol. 13, no. 6, pp. 739-750, 2004.
- [8] Shinyoung Lim, Kwanyong Lee, Okhwan Byeon and Taiyun Kim, 'Efficient Iris Recognition through Improvement of Feature Vector and Classifier', ETRI J., vol. 23, nNo. 2, pp. 61-70, 2001.
- [9] Lian Cai and Sidan Du, 'Rotation, scale and translation invariant image watermarking using Radon transform and Fourier transform', Proceedings of the IEEE 6th Circuit and systems Symposium Emerging

Technologies: Mobile and Wireless Communication, Shanghai, China, pp. 281-284, 2004.

- [10] Mitra Abhishek and Banerjee S., 'A Regular Algorithm For Real Time Radon and Inverse Radon Transform', Proceedings of IEEE Acoustics, Speech and Signal Processing (ICASSP), Montreal, Quebec, Canada, pp. v.105- v.108, 2004.
- [11] Kourosh Jafari-Kkouzani and Hamid Soltanian-Zadeh, 'Rotation-Invariant Multiresolution Texture analysis using Radon and Wavelet Transforms', IEEE Transactions on Image Processing, vol. 14, no. 6, pp. 783-795, 2005.
- [12] Jun Zhang, Xiyuan Zhou and Erke Mao, 'Image Object Recognition based on Radon Transform', Proc. of IEEE 5th World Congress on Intelligent Control and Automation, Hangzhou, China, pp. 4070-4074, 2004.
- [13] Haward L. Resnikoff and Raymond O. Wells , 'Wavelet Analysis-The Scalable Structure of Information', Springer-Verlag, New York (ISBN: 81-8128-226-4), 1998.
- [14] James S. Walker, 'A Primer on Wavelets and their Scientific Applications', CRC Press LLC, USA, 1999.
- [15] Phil Picton, 'Introduction to Neural Networks', The Macmillan Press Ltd., First edition, Great Britain (ISBN:0-333-61832-7), 1994.
- [16] Bremananth R., and Chitra A., 'A new approach for iris pattern analysis based on wavelet and HNN' , Journal of CSI, vol. 36, no.2, pp. 33-41 (ISSN: 0254-7813), 2006.
- [17] Bremananth R., Chitra A., 'Real-Time Image Orientation Detection and Recognition', International Conference on Signal and Image Processing (ICSIP), Dec. 2006, pp. 460-461.
- [18] Bremananth R., and Chitra A, 'Rotation Invariant Recognition of Iris', Journal of Systems Science and Engineering, Systems Society of India, vol.17, no.1, pp.69-78, 2008.
- [19] Bremananth R., *Ph.D. Dissertation*, Anna University, Chennai, India, 2008.

AUTHORS PROFILE



Mrs. M. Sankari received her B.Sc. and M.Sc. degrees in Computer science from Bharathidasan University, respectively. She has completed her Master of Philosophy degree in Computer science from Regional Engineering College, Trichy. Presently, she is a Head of the department of MCA at NIET and pursuing her doctorate degree in computer science at Avinashilingam University, Coimbatore, India. She has published various technical papers at IEEE conferences. Her field of research includes Computer vision, Pattern

recognition, Analysis of algorithms, Data structure, Computer graphics and multimedia.



Bremananth R received the B.Sc and M.Sc. degrees in Computer Science from Madurai Kamaraj and Bharathidasan University, respectively. He obtained M.Phil. degree in Computer Science and Engineering from Government college of Technology, Bharathiar University. He received his Ph.D. degree from Department of Computer Science and Engineering, PSG College of Technology, Anna University, Chennai, India. Presently, he is working as a Post-doctoral Research Fellow, at Nanyang Technological University, Singapore. He received the M N Saha Memorial award for the best application oriented paper in 2006 by Institute of Electronics and Telecommunication Engineers (IETE). His fields of research are acoustic imaging, pattern recognition, computer vision, image processing, biometrics, multimedia and soft computing. Dr. Bremananth is a member of Indian society of technical education (ISTE), advanced computing society (ACS), International Association of Computer Science and Information Technology (IACIT) and IETE.

Empirical Mode Decomposition Analysis of Heart Rate Variability

C.Santhi.M.E., Assistant Professor,

Electronics and Communication Engineering, Government College of Technology, Coimbatore-641 013

N.Kumaravel Ph.D

Professor, Head of the Department,

Electronics and Communication Engineering, Anna University, Chennai-600 025.

Abstract

The analysis of heart rate variability (HRV) demands specific capabilities not provided by either parametric or nonparametric spectral estimation methods. Empirical mode decomposition (EMD) has the possibility of dealing with nonstationary and nonlinear embedded phenomena, for a proper assessment of dynamic and transient changes in amplitude and time scales of HRV signal. In this work EMD and a non-linear curve fitting technique are used to study half an hour HRV signal and its intrinsic mode function obtained from 20 healthy young control subjects, 20 healthy old control subjects and 20 subjects with long term ST. The intrinsic oscillations are measured by means of its meanperiod and variance. Significant meanperiod reduction is observed in the intrinsic time scales of healthy old control subjects and subjects with long term ST, which is used to classify the three groups of HRV signal with high sensitivity and specificity. The estimated slope using the non-linear curve fitting technique represents the flexibility of the cardiovascular system. The main advantage of this method is it does not make any prior assumption about the HRV signal being analyzed and no artificial information is introduced into the filtering method.

Index Terms- *Empirical Mode Decomposition, Heart Rate Variability, Intrinsic Mode Functions, RR intervals, nonlinear curve fitting.*

1. Introduction

Over the last 20 years there has been widespread interest in the study of variations in the beat-to-beat interval of heart known as heart rate variability (HRV) or RR interval variations. HRV has been used as a measure of mortality primarily with patients who have undergone cardiac surgery. Clinical depression strongly associated with mortality with such patients may be seen through a decrease in HRV [1]. HRV is a non invasive measure of autonomic nervous system balance. Heart rate is influenced by both sympathetic and parasympathetic (vagal) activities of ANS. The influence of both branches of the autonomic nervous system (ANS) is known as sympathovagal balance reflected in the RR interval changes. A low frequency (LF) component provides a measure of sympathetic effects on the heart and generally occurs in a band between 0.04 Hz and 0.15 Hz. A measurement of the influence of the vagus nerve in modulating the sinoatrial node can be made in the high frequency band (HF) loosely defined between 0.15 and 0.4 Hz known as respiratory sinus arrhythmia (RSA), and is a measure of cardiac parasympathetic activity. The ratio of

power in the LF and HF bands (LF/HF) provides the measure of cardiac sympathovagal balance. Empirical Mode Decomposition (EMD) retains the intrinsic nonlinear nonstationary property of the signal. Any intrinsic timescale derived from the signal is based on the local characteristics timescale of the signal [2-4]. EMD carries out layer upon layer sifting and obtains ordered array components from smallest scale (highest frequency) to largest scale (lowest frequency) [4]. Empirical mode decomposition has the possibility of dealing with nonstationary and nonlinear embedded phenomena, and owing to its suitability for a proper assessment of the dynamic and transient changes in amplitude and in frequency of the HRV components [2& 3].

Application of EMD to half an hour HRV data yields nine intrinsic mode functions (IMFs). The first scale represents the highest frequency or the shortest period component of the signal. The second scale represents the lower frequency or the longer period component of the signal. Similarly the last IMF represents the lowest time scale present in the HRV signal. The first two scales contain more than 85% of total signal power. The meanperiod and variance of IMFs are computed as time domain measures. The variance of IMF decreases exponentially with respect to increasing timescales (meanperiods). Using nonlinear curve fitting technique the IMFs variations are estimated. The estimated parameter represents the flexibility of the cardiovascular system. The methodology is applied to HRV signal obtained from 20 healthy young control subjects, 20 healthy old control subjects and 20 subjects with long term ST. The intrinsic time scale of IMF 2 classifies the three groups HRV signal with high sensitivity and specificity.

2. Empirical Mode Decomposition (EMD)

EMD is a procedure oriented adaptive method for decomposing non-linear non-stationary signals. The components resulting from EMD are called Intrinsic Mode Functions (IMFs) [2]. The IMFs are amplitude frequency modulated intrinsic signals. The IMF's represents the oscillatory modes imbedded in the signal. It should satisfies the following two conditions. 1. In the whole data set the number of extrema's and the number of zero crossings must be either equal or differ by at most one. 2. At any point the mean value of the envelope defined by the local minima and the envelope defined by the local maxima is zero.

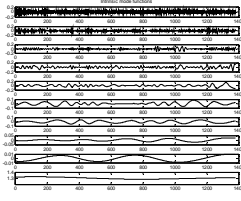
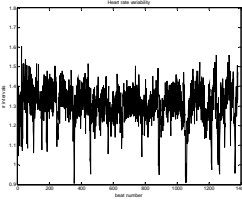


Fig. 1. RR interval signal

Fig. 2. Intrinsic Mode Functions

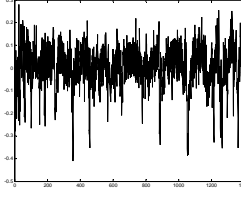
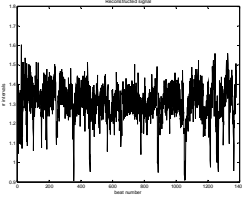


Fig. 3. Reconstructed signal

Fig. 4. Detrended signal

Figs 1-4 explain the efficiency of EMD for RR interval signal. The ECG data has been collected from the biomedical website [7] <http://www.physionet.org>. The RR intervals are derived from half an hour ECG signal by identifying the QRS complexes. The signal is manually edited and only noise free ectopic free segments are used for the analysis. A real time RR interval signal and its EMD decomposed IMFs are shown in Fig.1&2. Application of EMD to real time RR interval signal identifies eight to nine IMFs. The IMFs are zero mean amplitude frequency modulated signal. The decomposition is adaptive and lossless. The original RR interval signal is reconstructed using decomposed IMFs (Fig.3). The nonstationary trend is removed when the residue or monotonic trend (last IMF) is omitted while reconstructing the signal (Fig.4).

From the RR intervals the HRV signal or ΔRR signal ($R_{i+1}-R_i$) is obtained by computing successive difference between consecutive RR intervals. The obtained HRV signal and its IMFs are shown in Fig.5 and Fig.6. Matlab 7.1 tools are used for the analysis.

3. Methodology

SIFTING ALGORITHM:

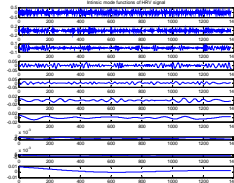
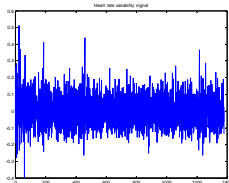


Fig. 5. HRV signal

Fig. 6. Intrinsic Mode Functions

Step 1: All the minima and maxima of the HRV Signal $x(t)$, are located.

Step 2: Spline Interpolate the minima and maxima points to obtain lower and upper envelopes of the signal.

Step 3: Compute mean envelope

$$m(t) = (\text{maxima's} + \text{minima's}) / 2.$$

Step 4: Subtract local mean from the original Signal to obtain local details $h(t) = x(t) - m(t)$.

Step 5: Check $h(t)$ for the conditions of an Intrinsic Mode Functions. [2]

If $h(t)$ is an IMF compute residue $r(t) = x(t) - h(t)$ and again the process is repeated to extract the next IMF. If $h(t)$ is not an IMF $x(t)$ is replaced with $h(t)$ and the procedure is repeated from step 1. Fig.6 shows all IMFs of the signal $x(t)$.

The process ends when the range of residue is below a predetermined level or the residue has a monotonic trend. In order to guarantee that the IMF components retains enough physical sense in both amplitude and frequency modulations, the sifting process is stopped by limiting the size of standard deviation (SD) which is computed from two consecutive sifting results.

$$SD = \sum_{t=0}^T \left[\left| h_{1(k-1)}(t) - h_{1k}(t) \right|^2 / h_{1(k-1)}^2(t) \right] \quad (1)$$

where k represents number of siftings.

The process of finding an intrinsic mode function requires number of iterations and the process to find all the IMFs requires further more iterations. As a result of this iterative procedure finally yields many IMFs and a residue. By summing up all the IMF functions and the residue, the original signal is reconstructed, given by the mathematical formulae

$$X(t) = \sum_{i=1}^n h_i(t) + r(n) \quad (2)$$

Where each h_i represents an intrinsic mode function and $r(n)$ either a mean trend or a constant.

For each IMF the meanperiod and variance are computed. The meanperiod is the ratio of distance between the first and last zero-crossings to number of zero-crossings of IMF.

The obtained RR interval signal using ECG represents the response of the cardiovascular system to ANS activities not the ANS activities themselves. The characteristics of cardiovascular system determine how the system responds to ANS activity and can alter significantly the characteristics of the HRV signal. The response characteristics are often nonlinear in nature. The IMFs capture the all the variations present in the HRV signal. Plotting the variance of all IMFs against its meanperiods gives a nonlinear function. The variance of IMF decreases with increasing meanperiod and this behavior is approximated using a geometric function

$$Y = aX^b \quad (3)$$

where Y represents vector of IMF's variance, X represents vector of meanperiods of IMFs, a is constant and b is the exponential decrease of the function. The IMFs meanperiod and variance of healthy young control subjects, healthy old control subjects and long term ST subjects vary significantly. The variations in the IMF are quantified by the parameter b . The parameter b represents the flexibility of cardiovascular system to ANS activities. The parameter b is estimated using nonlinear curve fitting technique explained below.

Taking logarithm of equation (3), gives

$$\ln Y = \ln a + b \ln X \quad (4)$$

putting $Y^* = \ln Y$, $X^* = \ln X$, $A^* = \ln a$ then the above nonlinear equation becomes $Y^* = A^* + bX^*$ which is a linear equation in X . The corresponding normal equations are

$$\sum Y^* = NA^* + b \sum X^* \quad (5)$$

$$\sum X^* Y^* = A^* \sum X^* + b \sum X^{*2} \quad (6)$$

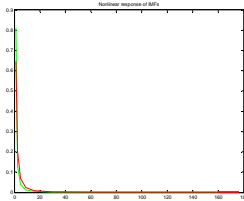


Fig.7. Curve fitting

N represents number of IMFs. Solving the normal equations using least mean square method the variables 'a' and 'b' are estimated [5]. The simulated response function using the estimated parameter is shown in Fig.7.

4. Results and Discussion

EMD and curve fitting techniques are applied to half an hour HRV signal of 20 healthy young control subjects, 20 healthy old control subjects and 20 subjects with long term ST. Empirical mode decomposition adaptively decomposes the half an hour HRV signal into number of Intrinsic mode functions (Fig.6). The analysis is done with ΔRR intervals. $\Delta RR (R_{i+1}-R_i)$ represents the difference between successive beat intervals. The IMFs are measured by their absolute variance, relative variance and meanperiods. The meanperiod is the ratio of distance between the first and last zero-crossings to the number of zero-crossings of IMF. First 3 IMFs contains more than 92% of total variance. The IMF1 represents the highest frequency or the shortest period component of the signal. The IMF2 contains the lower frequency or the longer period component of the signal. Since the 1st and 2nd IMF contains more than 85% of total power they are very significant.

Relative powers are computed with respect to total power considering all IMFs except the residue with zero meanperiod. In healthy young subjects an increase in relative power of IMF1 decreases the relative power of IMF2 (Fig.8). IMF 1 and IMF 2 are in phase opposition representing different components of the HRV signal. The original signal is interpolated to 2 Hz for a meaningful frequency measure. The Welch periodogram (with window width 1024 and window overlap of 512 samples) of IMFs of a healthy young control subject are shown in Fig.(9). Table-1 gives the peak frequency(Hz) and absolute spectralpower (ms^2 -milliseconds square) of IMFs The figure shows the frequency spectrum of the IMFs falls in the recognized spectral bands of HRV signal: 1.High frequency band from 0.15Hz to 0.5Hz; 2. Low frequency band from 0.04Hz to

0.15Hz; 3. Very low frequency band from 0.01Hz to 0.04Hz..

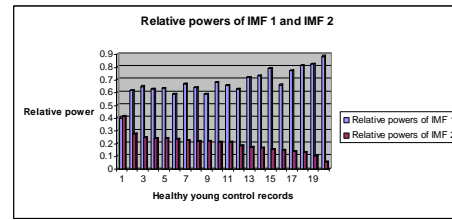


Fig 8: Relative powers of IMF 1 and IMF 2

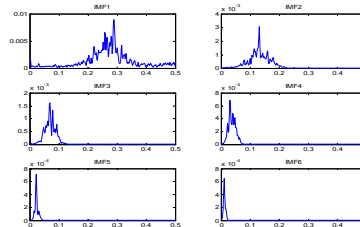


Fig 9: Welch periodogram of IMFs

IMFs	Peak frequency in Hz	Peak power in ms^2
IMF1	0.2891	0.01
IMF2	0.13	0.003
IMF3	0.068	0.002
IMF4	0.03	0.00069
IMF5	0.021	0.0007
IMF6	0.01	0.00062

Table-1 Spectral values of IMFs

The meanperiod of IMF2 of healthy young controls subjects are significantly higher compared to healthy old controls subjects and subjects with long term ST. Considering meanperiod of IMF2 (2.9724 secs) as threshold value, we classified the healthy young control subjects and subjects with long term ST with 95% sensitivity and 90% specificity. The classification is shown in Fig. (10). A threshold value of 2.8 secs classifies the healthy old controls subjects and subjects with long term ST with 90% sensitivity and 70% specificity shown in Fig .(11).

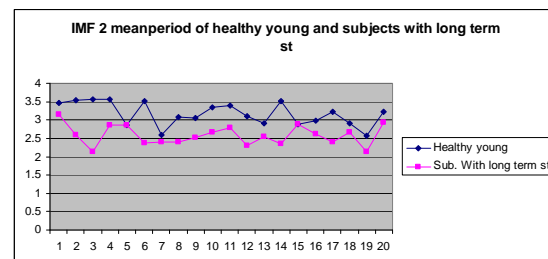


Fig.10. Meanperiod comparison of healthy young subjects and subjects with longterm ST.

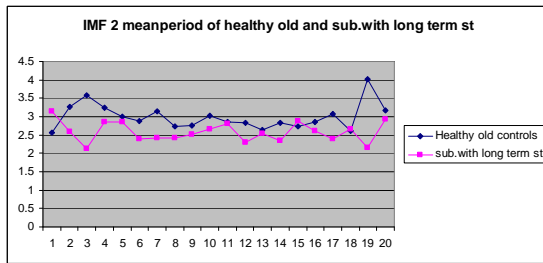


Fig.11. Meanperiod comparison of healthy old subjects and subjects with longterm ST.

The parameter b of IMFs of the three groups HRV signal are estimated, average plots are shown in Fig. (12). The estimated parameter b of healthy young control subjects, healthy old control subjects and long term ST subjects are -1.49, -1.43 and -1.39 (average values only). The more negative value represents the flexibility of the system. The healthy young control subject's cardiovascular system is more flexible than healthy old subjects and longterm ST subjects. The flexibility of the system decreases in healthy old control subjects and longterm ST subjects. The absolute powers of healthy young control subjects are significantly higher compared to healthy old subjects and long term ST subjects as shown in Fig.(13) (average values only). The higher values of absolute power represent more fluctuating power in the signal. The results show the HRV of healthy young control subjects contains higher power, longer time scales and more adaptive to ANS activities compared to healthy old control subjects and subjects with long term ST.

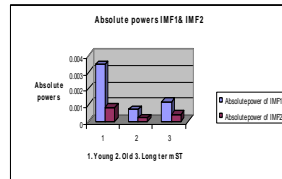
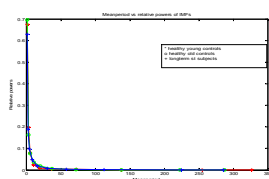


Fig.12 Correlation graphs Fig.13. Absolute powers

5. Conclusion

In order to cope up nonlinearity and nonstationarity issue of HRV signal EMD and nonlinear curve fitting techniques are used in this work. The IMFs of HRV signal are negatively correlated. The frequency spectrum of first two IMFs falls in the recognized HF and LF spectral bands of HRV signal. The meanperiod of IMF2 classifies half an hour HRV signal of healthy young control subjects, healthy old control subjects and subjects with long term ST with high sensitivity and specificity. The nonlinear curve fitting technique estimates the flexibility of cardiovascular system. The method is simple, adaptive and no artificial information is introduced in the analysis.

6. References

- [1] R. M. Carney, J. A. Blumenthal, P. K. Stein, L. Watkins, D. Catellier, L. F. Berkman, S. M. Czajkowski, C. O'Connor, P. H. Stone, K. E. Freedland, "Depression, Heart Rate Variability, and Acute Myocardial Infarction," *Circulation*, vol. 104, no. 17, pp. 2024 – 2028, 2001.
- [2] J.C.Echeverria, J.A.Crowe, M.S.Woolfson, B.R.Hayes-Gill, "Application of empirical mode decomposition to heart

rate variability analysis", *Med.Bio.Eng.Comput.*, 2001, 39, 471-479.

[3] E.P.Souza Neto, M.A.Custaud, J.C.Cejka, P.Abry, J.Frutoso, C.Gharib, P.Flandrin, "Assessment of Cardiovascular Autonomic Control by the Empirical Mode Decomposition", *Methods Inf Med* 2004;43:60-5.

[4] N.E.Huang, Z.Shen, S.R.Long, M.C.Wu, H.H.Shih, et al.1998. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis" *Proc.R.Soc.A*, vol 454, pp.903-995.

[5] B.V.Ramana, "Higher Engineering Mathematics", Tata McGraw-Hill Publishing Company Limited, New Delhi.

[6] HRV Analysis Software 1.1, developed by The Biomedical Signal Analysis Group, Department of Applied Physics, University of Kuopio, Finland. <http://venda.uku.fi/research/biosignal>

[7] www.physionet.org.

[8] Jan W.Kantelhard, Stephan A, Armin Bunde, 2002, Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series, *Physica A* 316, 87-114.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Mrs Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Mr. P. Vasant, University Technology Petronas, Malaysia
Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Mr. Praveen Ranjan Srivastava, BITS PILANI, India
Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia

Mr. Tirthankar Gayen, IIT Kharagpur, India
Ms. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Mr Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Mr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan
Prof. Syed S. Rizvi, University of Bridgeport, USA
Mr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Mr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Mr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Mr. S. Mehta, Inha University, Korea
Mr. Dilip Kumar S.M, University Visvesvaraya College of Engineering (UVCE), Bangalore University,
Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Mr. Saqib Saeed, University of Siegen, Germany
Mr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Mrs.J.Komala Lakshmi, SNR Sons College, Computer Science, India
Mr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, Faculty of Computer Science and Information Technology, University of
Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Mr. M. Azath, Anna University, India
Mr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Mr. Aos Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr Suresh Jain, Professor (on leave), Institute of Engineering & Technology, Devi Ahilya University, Indore
(MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Mr. Hanumanthappa. J. University of Mysore, India
Mr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Mr Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria

Mr. Santosh K. Pandey, Department of Information Technology, The Institute of Chartered Accountants of India

Dr. P. Vasant, Power Control Optimization, Malaysia

Dr. Petr Ivankov, Automatika - S, Russian Federation

Dr. Utkarsh Seetha, Data Infosys Limited, India

Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal

Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore

Assist. Prof. A. Neela madheswari, Anna university, India

Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India

Mr. Kamanashis Biswas, Daffodil International University, Bangladesh

Dr. Atul Gonsai, Saurashtra University, Gujarat, India

Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand

Mrs. G. Nalini Priya, Anna University, Chennai

Dr. P. Subashini, Avinashilingam University for Women, India

Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat

Mr Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal

Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India

Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof, Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India

Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah

Mr. Nitin Bhatia, DAV College, India

Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India

Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia

Assist. Prof. Sonal Chawla, Panjab University, India

Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India

Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia

Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia

Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India

Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France

Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India

Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology, Durban,South Africa

Prof. Mydhili K Nair, M S Ramaiah Institute of Technology(M.S.R.I.T), Affiliated to Visweswaraiah Technological University, Bangalore, India

M. Prabu, Adhiyamaan College of Engineering/Anna University, India

Mr. Swakkhar Shatabda, Department of Computer Science and Engineering, United International University, Bangladesh

Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan

Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India

Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India

Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Prof Ekta Walia Bhullar, Maharishi Markandeshwar University, Mullana (Ambala), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, : Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, College of Computer Science and Engineering, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institutiion of Engg. & Tech. CHD, India
Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Mr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India

Dr. C. Arun, Anna University, India

Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India

Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran

Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology

Subhabrata Barman, Haldia Institute of Technology, West Bengal

Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan

Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India

Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India

Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand

Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India

Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.

Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran

Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India

Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA

Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India

Dr. Umesh Kumar Singh, Vikram University, Ujjain, India

Mr. Serguei A. Mokhov, Concordia University, Canada

Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia

Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India

Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA

Dr. S. Karthik, SNS College of Technology, India

Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain

Mr. A.D.Potgantwar, Pune University, India

Dr. Himanshu Aggarwal, Punjabi University, India

Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India

Dr. K.L. Shunmuganathan, R.M.K Engg College, Kavaraipeitai, Chennai

Dr. Prasant Kumar Pattnaik, KIST, India.

Dr. Ch. Aswani Kumar, VIT University, India

Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA

Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan

Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia

Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA

Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India

Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India

Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia

Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA

Mr. R. Jagadeesh Kannan, RMK Engineering College, India

Mr. Deo Prakash, Shri Mata Vaishno Devi University, India

Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mahammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F.Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhanian University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S Balaram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A.Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praakash Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India

Dr . Mahendra Kumar , Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath , ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan
Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto , Singapore
Mr. Pawan Jindal, Jaypee University of Engineering and Technology, India
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhania University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhania University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India

Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India

Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India

Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India

Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India

Dr. Hanan Elazhary, Electronics Research Institute, Egypt

Dr. Hosam I. Faiq, USM, Malaysia

Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India

Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India

Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India

Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan

Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India.

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2010
ISSN: 1947-5500
<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, now at its sixth edition, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2010 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2010
ISSN 1947 5500